

Sequence 9, Application US/09126640A	Sequence 9, Application US/08944495-6	Sequence 9, Application US/08925588-6	Sequence 9, Application US/08925588-6
Patent No. 6099823	Patent No. 6087477	Patent No. 6221628	Patent No. 6221628
GENERAL INFORMATION:	GENERAL INFORMATION:	GENERAL INFORMATION:	GENERAL INFORMATION:
APPLICANT: FALB, Dean A.	APPLICANT: FALB, Dean	APPLICANT: FALB, Dean	APPLICANT: FALB, Dean
TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE	TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF	TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR	TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 7853-126	FILE REFERENCE: 7853-126	FILE REFERENCE: 7853-126	FILE REFERENCE: 7853-126
CURRENT APPLICATION NUMBER: US/09/126,640A	CURRENT APPLICATION NUMBER: US/08/944,495-6	CURRENT APPLICATION NUMBER: US/08/925,588	CURRENT APPLICATION NUMBER: US/08/925,588
CURRENT FILING DATE: 1998-07-30	CURRENT FILING DATE: 1998-07-30	CURRENT FILING DATE: 1998-07-30	CURRENT FILING DATE: 1998-07-30
EARLIER FILING DATE: 1997-06-06	EARLIER FILING DATE: 1997-06-06	EARLIER FILING DATE: 1997-06-06	EARLIER FILING DATE: 1997-06-06
EARLIER FILING DATE: 1997-02-13	EARLIER FILING DATE: 1997-02-13	EARLIER FILING DATE: 1997-02-13	EARLIER FILING DATE: 1997-02-13
EARLIER FILING DATE: 1996-02-16	EARLIER FILING DATE: 1996-02-16	EARLIER FILING DATE: 1996-02-16	EARLIER FILING DATE: 1996-02-16
NUMBER OF SEQ ID NOS: 44	NUMBER OF SEQ ID NOS: 44	NUMBER OF SEQ ID NOS: 44	NUMBER OF SEQ ID NOS: 44
SOFTWARE: FastSeq for Windows Version 4.0	SOFTWARE: FastSeq for Windows Version 4.0	SOFTWARE: FastSeq Version 2.0	SOFTWARE: FastSeq Version 2.0
SEQ ID NO 9	SEQ ID NO 9	SEQ ID NO 9	SEQ ID NO 9
LENGTH: 283	LENGTH: 283	LENGTH: 283	LENGTH: 283
TYPE: PRT	TYPE: PRT	TYPE: PRT	TYPE: PRT
ORGANISM: Homo sapiens	ORGANISM: Homo sapiens	ORGANISM: Homo sapiens	ORGANISM: Homo sapiens
US-09-126-640-9	US-09-126-640-9	US-09-126-640-9	US-09-126-640-9
Query Match	Query Match	Query Match	Query Match
Best Local Similarity 96.8%; Pred. No. 4.5e-15;	Best Local Similarity 96.8%; Pred. No. 4.5e-15;	Best Local Similarity 96.8%; Pred. No. 4.5e-15;	Best Local Similarity 96.8%; Pred. No. 4.5e-15;
Matches 30; Conservative 0; Indels 0; Gaps 0;	Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;	Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;	Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32	QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32	QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32	QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
DB 74 KWDNTLTGTEISWENKLAEGKLTLDITFV 104	DB 74 KWDNTLTGTEISWENKLAEGKLTLDITFV 104	DB 74 KWDNTLTGTEISWENKLAEGKLTLDITFV 104	DB 74 KWDNTLTGTEISWENKLAEGKLTLDITFV 104
RESULT 5	RESULT 5	RESULT 7	RESULT 7
US-08-944-495-6	US-08-944-495-6	US-08-925-588-6	US-08-925-588-6
Sequence 6, Application US/08944495	Sequence 6, Application US/08944495	Sequence 6, Application US/08925588	Sequence 6, Application US/08925588
Patent No. 6087477	Patent No. 6087477	Patent No. 6221628	Patent No. 6221628
GENERAL INFORMATION:	GENERAL INFORMATION:	GENERAL INFORMATION:	GENERAL INFORMATION:
APPLICANT: FALB, Dean	APPLICANT: FALB, Dean	APPLICANT: FALB, Dean	APPLICANT: FALB, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR	TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF	TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR	TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF	TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF	TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF	TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 44	NUMBER OF SEQUENCES: 44	NUMBER OF SEQUENCES: 44	NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:	CORRESPONDENCE ADDRESS:	CORRESPONDENCE ADDRESS:	CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP	ADDRESSEE: PENNIE & EDMONDS LLP	ADDRESSEE: PENNIE & EDMONDS LLP	ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas	STREET: 1155 Avenue of the Americas	STREET: 1155 Avenue of the Americas	STREET: 1155 Avenue of the Americas
CITY: New York	CITY: New York	CITY: New York	CITY: New York
STATE: NY	STATE: NY	STATE: NY	STATE: NY
COUNTRY: USA	COUNTRY: USA	COUNTRY: USA	COUNTRY: USA
ZIP: 10036-2711	ZIP: 10036-2711	ZIP: 10036-2711	ZIP: 10036-2711
COMPUTER READABLE FORM:	COMPUTER READABLE FORM:	COMPUTER READABLE FORM:	COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette	MEDIUM TYPE: Diskette	MEDIUM TYPE: Diskette	MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible	COMPUTER: IBM Compatible	COMPUTER: IBM Compatible	COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS	OPERATING SYSTEM: DOS	OPERATING SYSTEM: DOS	OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0	SOFTWARE: FastSeq Version 2.0	SOFTWARE: FastSeq Version 2.0	SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:	CURRENT APPLICATION DATA:	CURRENT APPLICATION DATA:	CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,495	APPLICATION NUMBER: US/08/944,495	APPLICATION NUMBER: US/08/925,588	APPLICATION NUMBER: US/08/925,588
FILING DATE:	FILING DATE:	FILING DATE:	FILING DATE:
CLASSIFICATION:	CLASSIFICATION:	CLASSIFICATION:	CLASSIFICATION:
PRIOR APPLICATION DATA:	PRIOR APPLICATION DATA:	PRIOR APPLICATION DATA:	PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910	APPLICATION NUMBER: 08/799,910	APPLICATION NUMBER: 08/799,910	APPLICATION NUMBER: 08/799,910
FILING DATE:	FILING DATE:	FILING DATE:	FILING DATE:
ATTORNEY/AGENT INFORMATION:	ATTORNEY/AGENT INFORMATION:	ATTORNEY/AGENT INFORMATION:	ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A	NAME: Coruzzi, Laura A	NAME: Coruzzi, Laura A	NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742	REGISTRATION NUMBER: 30,742	REGISTRATION NUMBER: 30,742	REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-067-999	REFERENCE/DOCKET NUMBER: 7853-067-999	REFERENCE/DOCKET NUMBER: 7853-067-999	REFERENCE/DOCKET NUMBER: 7853-067-999
TELEPHONE: (212)7909090	TELEPHONE: (212)7909090	TELEPHONE: (212)7909090	TELEPHONE: (212)7909090
TELEFAX: (212)8699741	TELEFAX: (212)8699741	TELEFAX: (212)8699741	TELEFAX: (212)8699741
INFORMATION FOR SEQ ID NO: 6:	INFORMATION FOR SEQ ID NO: 6:	INFORMATION FOR SEQ ID NO: 6:	INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:	SEQUENCE CHARACTERISTICS:	SEQUENCE CHARACTERISTICS:	SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids	LENGTH: 283 amino acids	LENGTH: 283 amino acids	LENGTH: 283 amino acids
TYPE: amino acid	TYPE: amino acid	TYPE: amino acid	

STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-08-925-588-6

Query Match 72.4%; Score 160; DB 3; Length 283;
Best Local Similarity 96.8%; Pred. No. 4.5e-15;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
|||:|||||
Db 74 KWTNTLTGTEISWENKLAEGKLTLDITFV 104
|||:|||||

RESULT 8
US-09-288-292A-9
; Sequence 9, Application US/09288292A
; Patent No. 6359194
; GENERAL INFORMATION:
; APPLICANT: Dean A. Falb
; APPLICANT: Katherine Galvin
; APPLICANT: Michael Donovan
; APPLICANT: Dennis Huszar
; APPLICANT: Michael A. Gimbrone, Jr.
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Diagnosis of
; FILE REFERENCE: 7853-140-999
; CURRENT APPLICATION NUMBER: US/09/288,292A
; CURRENT FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 08/870,434
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
; PRIOR FILING DATE: 1996-02-16
; PRIOR APPLICATION NUMBER: 08/485,573
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/386,844
; PRIOR FILING DATE: 1995-02-10
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-288-292A-9

Query Match 72.4%; Score 160; DB 3; Length 283;
Best Local Similarity 96.8%; Pred. No. 4.5e-15;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
|||:|||||
Db 74 KWTNTLTGTEISWENKLAEGKLTLDITFV 104
|||:|||||

RESULT 9
US-09-372-044-6
; Sequence 6, Application US/09372044A
; Patent No. 6492126
; GENERAL INFORMATION:
; APPLICANT: Dean Falb et al.
; TITLE OF INVENTION: Compositions and Methods for the
; FILE REFERENCE: 7853-152
; CURRENT APPLICATION NUMBER: US/09/372,044A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 283

TYPE: PRT
ORGANISM: Homo sapiens
US-09-372-044-6

Query Match 72.4%; Score 160; DB 4; Length 283;
Best Local Similarity 96.8%; Pred. No. 4.5e-15;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
|||:|||||
Db 74 KWTNTLTGTEISWENKLAEGKLTLDITFV 104
|||:|||||

RESULT 10
US-08-825-486-6
; Sequence 6, Application US/08825486
; Patent No. 6534641
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/825,486
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-077-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 7909090
; TELEFAX: (212) 8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-825-486-6

Query Match 72.4%; Score 160; DB 4; Length 283;
Best Local Similarity 96.8%; Pred. No. 4.5e-15;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
|||:|||||
Db 74 KWTNTLTGTEISWENKLAEGKLTLDITFV 104
|||:|||||

RESULT 11
US-08-826-248-6
; Sequence 6, Application US/08826248
; Patent No. 6759210

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,227
; FILING DATE: Filed herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0135 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 340201
;
US-08-726-227-3
;
Query Match 50.7%; Score 112; DB 1; Length 292;
Best Local Similarity 66.7%; Pred No. 4,1e-08;
Matches 20; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 KWDNTNLGTGISWENKLAEGKLTLDTIF 31
   |||||||||:|:::|||||||
DbB 84 KWNTDNLGTGIAIEDQICQGLKLTFTF 113

RESULT 13
US-09-489-039A-13424
; Sequence 13424, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13424
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13424
;
Query Match 24.0%; Score 53; DB 4; Length 383;
Best Local Similarity 34.2%; Pred. No. 19;
Matches 13; Conservative 4; Mismatches 15; Indels 6; Gaps 1;

QY 3 WDTDNLGTGISWENKLAEGKLTLDTIFVHHVLAHP 40
   |||||:|:|:|:|:|:|
DbB 213 WGMDCALITEADWQAVPASDELTV-----HVLPAH 244

RESULT 14
US-09-886-319A-64

```

Job time : 12.9136 secs

```
; Sequence 64, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Coppel, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886,319A
; CURRENT FILING DATE: 2001-05-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-319A-64
```

```
Query Match 24.0%; Score 53; DB 4; Length 822;
Best Local Similarity 37.0%; Pred. No. 49;
Matches 10; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 4 DTDNLTGTEISWENKLAEGKLTLDTI 30
: : ||| ||| ||| ||| :
Db 286 ENENLQANEIMWNNLTAEISLQVLMKTL 312
```

```
RESULT 15
US-09-538-092-941
; Sequence 941, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 941
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P16591
US-09-538-092-941
```

```
Query Match 24.0%; Score 53; DB 4; Length 822;
Best Local Similarity 37.0%; Pred. No. 49;
Matches 10; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 4 DTDNLTGTEISWENKLAEGKLTLDTI 30
: : ||| ||| ||| ||| :
Db 286 ENENLQANEIMWNNLTAEISLQVLMKTL 312
```


APPLICANT: Agee et al.
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-502C
CURRENT APPLICATION NUMBER: US/10/307,817
CURRENT FILING DATE: 2002-12-02
NUMBER OF SEQ ID NOS: 682
SOFTWARE: CuraseqList version 0.1
SEQ ID NO 140
LENGTH: 247
TYPE: PRT
ORGANISM: Homo sapiens
US-10-307-817-140

Query Match 72.4%; Score 160; DB 15; Length 247;
Best Local Similarity 96.8%; Pred. No. 4.8e-14;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDTFV 32
||:|||||
Db 75 KWDNTLTGTEISWENKLAEGKLTLDTFV 105

RESULT 3
US-08-825-486-6
Sequence 6, Application US/08825486
Publication No. US20020016303A1
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
THE TREATMENT AND DIAGNOSIS OF
TITLE OF INVENTION: CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,486
FILING DATE: 28-MAR-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-077-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)7909090
TELEFAX: (212)8699741
TELEX: 66141 PENNIE

SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-825-486-6

Query Match 72.4%; Score 160; DB 8; Length 283;
Best Local Similarity 96.8%; Pred. No. 4.8e-14;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDTFV 32
||:|||||
Db 74 KWDNTLTGTEISWENKLAEGKLTLDTFV 104

RESULT 4
US-08-870-434-9
Sequence 9, Application US/08870434
Publication No. US20020034736A1
GENERAL INFORMATION:
APPLICANT: Falb, Dean
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,434
FILING DATE: 06-JUN-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/799,910
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-084
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-870-434-9

Query Match 72.4%; Score 160; DB 8; Length 283;
Best Local Similarity 96.8%; Pred. No. 4.8e-14;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDTFV 32
||:|||||
Db 74 KWDNTLTGTEISWENKLAEGKLTLDTFV 104

RESULT 5
US-09-372-044-6
Sequence 6, Application US/09372044A
Patent No. US20020102603A1
GENERAL INFORMATION:
APPLICANT: Dean Falb et al.
TITLE OF INVENTION: Compositions and Methods for the
Treatment and Diagnosis of Cardiovascular Disease
FILE REFERENCE: 7853-152
CURRENT APPLICATION NUMBER: US/09/372,044A
CURRENT FILING DATE: 1999-08-11
NUMBER OF SEQ ID NOS: 44
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6

```
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-372-044-6

Query Match
Best Local Similarity 72.4%; Score 160; DB 9; Length 283;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWTDTNTLGTISWENKLAEGKLTLDITFV 32
Db 74 KWTDTNTLGTISWENKLAEGKLTLDITFV 104

RESULT 6
US-09-560-150-9
; Sequence 9, Application US/09560150
; Publication No. US20030073076A1
; GENERAL INFORMATION:
; APPLICANT: FALB, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/560,150
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 09/126,640
; PRIOR FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: 08/870,434
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
; PRIOR FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-560-150-9

Query Match
Best Local Similarity 72.4%; Score 160; DB 10; Length 283;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWTDTNTLGTISWENKLAEGKLTLDITFV 32
Db 74 KWTDTNTLGTISWENKLAEGKLTLDITFV 104

RESULT 7
US-10-067-741-9
; Sequence 9, Application US/10067741
; Publication No. US20030097668A1
; GENERAL INFORMATION:
; APPLICANT: Dean A. Falb
; APPLICANT: Katherine Galvin
; APPLICANT: Michael Donovan
; APPLICANT: Dennis Huszar
; APPLICANT: Michael A. Gimbrone, Jr.
; TITLE OF INVENTION: Compositions and Methods for the Treatment and
; Diagnosis of
; TITLE OF INVENTION: Cardiovascular Disease
; FILE REFERENCE: 7853-140-999
; CURRENT APPLICATION NUMBER: US/10/067,741
; CURRENT FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US/09/288,292
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 08/870,434
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
```

```
; PRIOR FILING DATE: 1996-02-16
; PRIOR APPLICATION NUMBER: 08/485,573
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/386,844
; PRIOR FILING DATE: 1995-02-10
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-067-741-9

Query Match
Best Local Similarity 72.4%; Score 160; DB 14; Length 283;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWTDTNTLGTISWENKLAEGKLTLDITFV 32
Db 74 KWTDTNTLGTISWENKLAEGKLTLDITFV 104

RESULT 8
US-10-264-049-2654
; Sequence 2654, Application US/10264049
; Publication No. US20040008579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2654
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-264-049-2654

Query Match
Best Local Similarity 72.4%; Score 160; DB 15; Length 284;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWTDTNTLGTISWENKLAEGKLTLDITFV 32
Db 75 KWTDTNTLGTISWENKLAEGKLTLDITFV 105

RESULT 9
US-10-408-765A-1427
; Sequence 1427, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Falty, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1427
```

```

; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1427

Query Match          72.4%; Score 160; DB 16; Length 284;
Best Local Similarity 96.8%; Pred. No. 4.8e-14;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
DB 75 KWTDTLTGTEISWENKLAEGKLTLDITFV 105

RESULT 10
US-10-408-765A-3005
; Sequence 3005, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary W.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3005
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-3005

Query Match          72.4%; Score 160; DB 16; Length 284;
Best Local Similarity 96.8%; Pred. No. 4.8e-14;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
DB 75 KWTDTLTGTEISWENKLAEGKLTLDITFV 105

RESULT 11
US-09-930-512-14
; Sequence 14, Application US/09930512
; Publication No. US20040010118A1
; GENERAL INFORMATION:
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spytek, Kimberly
; APPLICANT: Spaderna, Steven
; APPLICANT: Gangolli, Esha A
; APPLICANT: Rastelli, Luca
; APPLICANT: Burgess, Catherine E
; APPLICANT: Majumder, Kunud
; APPLICANT: Shimkets, Richard
; APPLICANT: Mishra, Vishnu
; APPLICANT: Vernet, Corine
; APPLICANT: Szekeres, Edward S
; APPLICANT: Grosse, William M
; APPLICANT: Alsbrook II, John P
; APPLICANT: Liu, Xiaohong
; APPLICANT: Gerlach, Valerie L
; APPLICANT: Ellerman, Karen
; APPLICANT: Smithson, Glenda
; APPLICANT: Peyman, John
; APPLICANT: Stone, David
US-09-930-512-14

Query Match          72.4%; Score 160; DB 16; Length 284;
Best Local Similarity 96.8%; Pred. No. 4.8e-14;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
DB 75 KWTDTLTGTEISWENKLAEGKLTLDITFV 105

RESULT 12
US-10-408-765A-1647
; Sequence 1647, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary W.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1647
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1647

Query Match          50.7%; Score 112; DB 16; Length 293;
Best Local Similarity 66.7%; Pred. No. 3e-07;
Matches 20; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 31
DB 74 KWTDTLTGTEISWENKLAEGKLTLDITFV 104

Query Match          53.8%; Score 119; DB 11; Length 283;
Best Local Similarity 80.6%; Pred. No. 3e-08;
Matches 25; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
DB 74 KWTDTLTGTEISWENKLAEGKLTLDITFV 104

US-09-930-512-14
; ORGANISM: Homo sapiens
; TYPE: PRT
; LENGTH: 283
; SEQ ID NO 14
; SOFTWARE: Patentin Ver. 2.1
; ORGANISM: Homo sapiens
US-09-930-512-14

Query Match          53.8%; Score 119; DB 11; Length 283;
Best Local Similarity 80.6%; Pred. No. 3e-08;
Matches 25; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
DB 74 KWTDTLTGTEISWENKLAEGKLTLDITFV 104

US-10-408-765A-1647
; Sequence 1647, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary W.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1647
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1647

Query Match          50.7%; Score 112; DB 16; Length 293;
Best Local Similarity 66.7%; Pred. No. 3e-07;
Matches 20; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 31
DB 74 KWTDTLTGTEISWENKLAEGKLTLDITFV 104

```

```

; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Piroock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018201
; CURRENT APPLICATION NUMBER: US/10/205,194
; CURRENT FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Voltage-dependent anion channel 2
US-10-205-194-17

Query Match      50.7%; Score 112; DB 14; Length 295;
Best Local Similarity 66.7%; Pred. No. 3.1e-07;
Matches 20; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy      2 KWDTDNLGTEISWENKLAEGKLTLDTTF 31
Db      86 KWTNDTLGTEIAIEDQICQGLKLTFTD 115

Search completed: November 10, 2004, 16:35:48
Job time : 39.5802 secs

; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 108
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 900341CD1
US-09-981-353-108

Query Match      50.7%; Score 112; DB 9; Length 294;
Best Local Similarity 66.7%; Pred. No. 3e-07;
Matches 20; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy      2 KWDTDNLGTEISWENKLAEGKLTLDTTF 31
Db      85 KWTNDTLGTEIAIEDQICQGLKLTFTD 114

RESULT 14
US-10-408-765A-673
; Sequence 673, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 860088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 673
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-673

Query Match      50.7%; Score 112; DB 16; Length 294;
Best Local Similarity 66.7%; Pred. No. 3e-07;
Matches 20; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy      2 KWDTDNLGTEISWENKLAEGKLTLDTTF 31
Db      85 KWTNDTLGTEIAIEDQICQGLKLTFTD 114

RESULT 15
US-10-205-194-17
; Sequence 17, Application US/10205194
; Publication No. US20030134301A1
; GENERAL INFORMATION:

```

This Page Blank (uspio)

QY 2 KWDTDNTLGTISWENKLAEGKLTLDITF 31
 DB 86 KWDTDNTLGTISWENKLAEGKLTLDITF 115

RESULT 3
 A45972
 Mitochondrial porin, long form - human
 C:Species: Homo sapiens (man)
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
 C:Accession: A45972
 R:Ha H.; Hajek, P.; Bedwell, D.M.; Burrows, P.D.
 J. Biol. Chem. 268, 12143-12149, 1993
 A:Title: A mitochondrial porin cDNA predicts the existence of multiple human porins.
 A:Reference number: A45972; MUID:93280191; PMID:7685033
 A:Accession: A45972
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-347 <HAA>
 A:Cross-references: UNIPROT:P45880; GB:L08666; NID:G190199; PIDN:AAA60144.1; PID:G190200
 C:Superfamily: porin
 C:Keywords: alternative initiators; mitochondrion

Query Match 50.7%; Score 112; DB 2; Length 347;
 Best Local Similarity 66.7%; Pred. No. 2.9e-07;
 Matches 20; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 KWDTDNTLGTISWENKLAEGKLTLDITF 31
 DB 100 KWDTDNTLGTISWENKLAEGKLTLDITF 129

RESULT 4
 A36875
 Plasmalemmal voltage-dependent anion channel protein 1 - bovine (fragment)
 N:Alternate names: plasmalemmal porin
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
 C:Accession: A36875; S40132
 R:Dermietzel, R.; Hwang, T.K.; Buettner, R.; Hofer, A.; Dotzler, E.; Kremer, M.; Deutzma
 Proc. Natl. Acad. Sci. U.S.A. 91, 499-503, 1994
 A:Title: Cloning and in situ localization of a brain-derived porin that constitutes a la
 A:Reference number: A36875; MUID:94119914; PMID:7507248
 A:Accession: A36875
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-282 <DER>
 A:Cross-references: UNIPROT:P45879; GB:X75068; NID:G437026; PIDN:CAA52962.1; PID:G437027
 C:Superfamily: porin
 C:Keywords: voltage-gated ion channel

Query Match 50.2%; Score 111; DB 2; Length 282;
 Best Local Similarity 70.0%; Pred. No. 3e-07;
 Matches 21; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 KWDTDNTLGTISWENKLAEGKLTLDITF 31
 DB 73 KWDTDNTLGTISWENKLAEGKLTLDITF 102

RESULT 5
 M45973
 voltage-dependent anion channel 1 [validated] - human
 N:Alternate names: 31K porin, lymphocyte; Mitochondrial porin 1; porin 31HL; porin 31HM
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1991 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004
 C:Accession: A44422; S07478; S16195; S04018
 R:Blachly-Dyson, E.; Zamboni, E.B.; Yu, W.H.; Adams, V.; McCabe, E.R.; Adelman, J.; C
 J. Biol. Chem. 268, 1835-1841, 1993
 A:Title: Cloning and functional expression in yeast of two human isoforms of the outer m
 A:Reference number: A44422; MUID:93131931; PMID:8420959
 A:Accession: A44422

A:Molecule type: mRNA
 A:Residues: 1-283 <BLA>
 A:Cross-references: UNIPROT:P21796; GB:L06132; NID:G340198; PIDN:AAA61272.1; PID:G340199
 A:Experimental source: B cell line WIL-2NS
 A>Note: sequence extracted from NCBI backbone (NCBI:122926)
 R:Kayser, H.; Kratzin, H.D.; Thinnies, F.P.; Goetz, H.; Schmidt, W.E.; Eckart, K.; Hilsch
 Biol. Chem. Hoppe-Seyler 370, 1265-1278, 1989
 A:Title: To the knowledge of human porins. II. Characterization and primary structure of
 A:Reference number: S07478; MUID:90148194; PMID:2559745
 A:Accession: S07478
 A:Molecule type: protein
 A:Residues: 2-283 <KAY>
 A>Note: article in German with English abstract
 R:Jurgens, L.; Ilse, H.; Kratzin, H.D.; Hesse, D.; Eckart, K.; Thinnies, F.P.; Hilsch
 Biol. Chem. Hoppe-Seyler 372, 455-463, 1991
 A:Title: Studies on human porin. IV. The primary structures of "Porin 31HM" purified from
 A:Reference number: S16195; MUID:92029673; PMID:1657034
 A:Accession: S16195
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 2-283 <RPH>
 R:Kayser, H.; Kratzin, H.D.; Thinnies, F.P.; Goetz, H.; Hilschmann, N.
 submitted to the Protein Sequence Database, April 1989
 A:Reference number: S04018
 A:Molecule type: protein
 A:Residues: 2-283 <KAT>
 C:Genetics:
 A:Gene: GDB:VDAC1
 A:Cross-references: GDB:138280; OMIM:314555
 A:Map position: Xq13-Xq21
 C:Superfamily: porin
 C:Keywords: acetylated amino end; mitochondrial outer membrane
 F:2-283/Product: voltage-dependent anion channel 1 #status experimental <MAT>
 F:2/Modified site: acetylated amino end (Ala) (in mature form) #status experimental

Query Match 50.2%; Score 111; DB 1; Length 283;
 Best Local Similarity 70.0%; Pred. No. 3.1e-07;
 Matches 21; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 KWDTDNTLGTISWENKLAEGKLTLDITF 31
 DB 74 KWDTDNTLGTISWENKLAEGKLTLDITF 103

RESULT 6
 S65537
 porin - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
 C:Accession: S65537; S65506
 R:Messina, A.; Neri, M.; Perosa, F.; Caggese, C.; Marino, M.; Caizzi, R.; de Pinto, V.
 FEBS Lett. 384, 9-13, 1996
 A:Title: Cloning and chromosomal localization of a cDNA encoding a mitochondrial porin fi
 A:Reference number: S65506; MUID:96390820; PMID:8797793
 A:Accession: S65537
 A:Molecule type: mRNA
 A:Residues: 1-280 <MES>
 A:Cross-references: EMBL:X92408
 A:Accession: S65506
 A:Molecule type: protein
 A:Residues: 2-10 <NEW>
 C:Genetics:
 A:Genome: nuclear
 C:Superfamily: porin
 C:Keywords: mitochondrion
 F:2-280/Product: porin #status experimental <MAT>

Query Match 40.3%; Score 89; DB 2; Length 280;
 Best Local Similarity 59.3%; Pred. No. 0.00028;
 Matches 16; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 2 KWDTDNTLGTISWENKLAEGKLTLDITF 28

Df
73 KWTNDTLFTFEVAVQDQLLEGLKLSUE 99

|||||
||:::||:
||:::||:

hypothetical protein R05G6.7 - Caenorhabditis elegans
C;Species:Caenorhabditis elegans
C;Date:15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C;Accession:T29355
R;Murray, J.; Le, T.T.
submitted to the EMBL Data Library, May 1996
A;Description:The sequence of C. elegans cosmid R05G6.
A;Reference number: Z20612
A;Accession:T29355
A;Status:Preliminary; translated from GB/EMBL/DDBU
A;Molecule type: DNA
A;Residues: 1-310 <MR>
A;Cross-references: EMBL:U58746; PIDN:AAB00622.1; GSPDB:GN00022; CESP:R05G6.7
A;Experimental source: strain Bristol N2; clone R05G6
C;Genetics:
A;Gene: CESP:R05G6.7
A;Map position: 4
A;Introns: 22/1; 69/1; 152/3; 238/3
C;Superfamily: porin

Query Match 39.8% Score 88; DB 2; Length 310;
Best Local Similarity 46.2%; Pred.No. 0.00043;
Matches 18; Conservative 6; Mismatches 9; Indels 6; Gaps 1;

Qy 2 KWTDNTLTGTISWENKLAEGLKLTLDITFVVHVLAPH 40
||:::||:
||:::||:

Df 75 KWNTENGTVIEVNEGFGRGLKVTLDS-----LYAPH 107
|||||
||:::||:

cinnamyl-alcohol dehydrogenase (EC 1.1.1.195) F28A23.10 - Arabidopsis thaliana
C;Species:Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T05413
R;Bevan, M.; Weichselgartner, M.; Fatmann, B.; Granderath, X.; Dauner, D.; Herzl, A.; Menz,
submitted to The Protein Sequence Database, October 1998
A;Reference number: Z15415
A;Accession: TC05413
A;Molecule type: DNA
A;Residues: 1-357 <BEV>
A;Cross-references: UNIPROT:O49482; EMBL:AL021961
A;Experimental source: cultivar Columbia; BAC clone F28A23
C;Genetics:
A;Map position: 4
A;Introns: 30/2; 68/2; 144/2; 291/1
A;Note: F28A23.10
C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
F;Keywords: metalloprotein; oxidoreductase; zinc
F;32-340/Domain: long-chain alcohol dehydrogenase homology<IAD>
F;183-212/Region: beta-alpha-beta NADP nucleotide-binding fold
F;47,69,163/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F;100,103,106,114/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 26.2% Score 58; DB 2; Length 357;
Best Local Similarity 35.5%; Pred.No. 5.5;
Matches 11; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

Qy 6 DNLTGTSIWENKLAEGLKLTDTIFVHHVL 36
||:::||:
||:::||:

Df 228 DYVGSDQAQMSELADSLDVVDTPVPVHAL 258
|||||
||:::||:

S7399S
probable lipoprotein VxpSP7_orf320 - Mycoplasma pneumoniae (strain ATCC 29342)

N;Alternate names: MG149 homolog VxpSP7 orf320

Job time : 10.0864 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: November 10, 2004, 11:27:01 ; Search time 43.7654 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-12
Perfect score: 221
Sequence: 1 MKWDTNTLTGTEISWENKLAEGKLTLDTIFVHVLHAPH 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues
Total number of hits satisfying chosen parameters: 1825181
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match %	Length	ID	Description	
1	160	72.4	78	2	Q9GL29	bos taurus
2	160	72.4	283	1	POR3_BOVIN	
3	160	72.4	283	1	POR3_HUMAN	
4	160	72.4	283	1	POR3_MOUSE	
5	160	72.4	284	2	Q8BNG2	mus musculus
6	157	71.0	283	1	POR3_RAT	
7	157	71.0	283	2	AAH6I780	rattus norvegicus
8	147	66.5	283	1	POR3_RABIT	
9	147	66.5	283	2	Q9MZ14	sus scrofa
10	141	63.8	112	1	POR3_PIG	
11	125	56.6	283	2	Q6GR11	xenopus laevis
12	115	52.0	282	1	POR2_MELGA	
13	115	52.0	282	1	POR2_XENLA	
14	115	52.0	283	2	Q8AWD0	brachydanio rerio
15	115	52.0	283	2	Q919D1	gallus gallus
16	115	52.0	283	2	AAQ94582	brachydanio rerio
17	115	52.0	283	2	AAH62525	brachydanio rerio
18	112	50.7	283	2	Q7L3U5	homo sapiens
19	112	50.7	283	2	Q7ZWZ0	xenopus laevis
20	112	50.7	283	2	CAG33245	homo sapiens
21	112	50.7	294	1	POR2_RABIT	
22	112	50.7	294	2	Q9BWK8	homo sapiens
23	112	50.7	294	2	Q9MYV7	bos taurus
24	112	50.7	294	2	Q9MZ15	sus scrofa
25	112	50.7	294	2	AAH00165	homo sapiens
26	112	50.7	294	2	AAH72407	homo sapiens
27	112	50.7	295	1	POR2_MOUSE	
28	112	50.7	295	1	POR2_RAT	
29	112	50.7	295	2	Q9L158	rattus norvegicus
30	112	50.7	295	2	AAQ01516	mus musculus
31	112	50.7	347	1	POR2_HUMAN	homo sapiens

ALIGNMENTS

RESULT 1			
Q9GL29	PRELIMINARY;	PRT;	78 AA.
ID	Q9GL29		
AC	Q9GL29;		
DT	01-MAR-2001 (T-EMBLrel. 16, Created)		
DT	01-MAR-2001 (T-EMBLrel. 16, Last sequence update)		
DT	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)		
DE	Voltage-dependent anion channel 3 (Fragment).		
GN	Name=VDAC3;		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Testis;		
RA	Asmarinah, Konrad L., Hirsch E., Schill W-B., Hirsch K-D.;		
RT	"Expression of porins in the testis.";		
RL	Andrologia 33:235-237(2001).		
DR	EMBL; AJ299423; CAC14092.1; --		
DR	GO; GO:0005741; C:Mitochondrial outer membrane; IEA.		
DR	GO; GO:0008308; F:voltage-dependent ion-selective channel act. .; IEA.		
DR	GO; GO:0006820; P:anion transport; IEA.		
DR	InterPro; IPR001925; Porin_Euk.		
DR	Pfam; PF01459; Porin_3; 1.		
FT	NON TER 1		
FT	NON TER 78		
SQ	SEQUENCE 78 AA; 8679 MW; 389D8F04D92D50ED CRC64;		
Query Match 72.4%; Score 160; DB 2; Length 78;			
Best Local Similarity 96.8%; Pred.No.1.6e-13;			
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;			
Qy	2 KWDNTNTLTGTEISWENKLAEGKLTLDTIFV 32		
Db	30 KWDNTNTLTGTEISWENKLAEGKLTLDTIFV 60		
RESULT 2			
POR3_BOVIN	STANDARD;	PRT;	283 AA.
ID	POR3_BOVIN		
AC	Q9MZ13;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	05-JUN-2004 (Rel. 44, Last annotation update)		
DE	Voltage-dependent anion-selective channel protein 3 (VDAC-3) (hVDAC3)		
DE	(Outer mitochondrial membrane protein porin 3).		
GN	Name=VDAC3;		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovinae; Bos.		
OX	NCBI_TaxID=9913;		

RN SEQUENCE FROM N.A.
 RC TISSUE=Lens;
 RA J.L.;
 RA "Ion channels in the lens."; EMBL/GenBank/DBJ databases.
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Forms a channel through the mitochondrial outer membrane
 CC that allows diffusion of small hydrophilic molecules (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Outer mitochondrial membrane.
 CC -!- DOMAIN: Consists mainly of membrane-spanning sided beta-sheets.
 CC -!- SIMILARITY: Belongs to the eukaryotic mitochondrial porin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AF268466; AAF80103.1; -.
 CC InterPro; IPR001925; Porin_Euk.
 CC Pfam; PF01459; Porin_3; 1.
 CC PRINTS; PR00185; EUKARYOTIC_PORIN.
 CC PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
 KW Mitochondrion; Outer membrane; Porin.
 SQ SEQUENCE 283 AA; 30739 MW; D305DA2EB42BEC3B CRC64;

 Query Match 72.4%; Score 160; DB 1; Length 283;
 Best Local Similarity 96.8%; Pred. No. 6.7e-13;
 Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 QY 2 KWDNTLGTISWENKLAEGKLTLDTFV 32
 ||:|||||
 DB 74 KWDNTLGTISWENKLAEGKLTLDTFV 104

 RESULT 3
 POR3 HUMAN
 ID POR3 HUMAN STANDARD; PRT; 283 AA.
 AC Q9Y277; Q9UISO;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Voltage-dependent anion-selective channel protein 3 (VDAC-3) (hVDAC3)
 DE (Outer mitochondrial membrane protein porin 3).
 GN Name=VDAC3;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Umbilical cord blood;
 RX MEDLINE=98318631; PubMed=9853160;
 RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
 RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
 RA Wang Y.-X., Chen S.-J., Chen Z.;
 RT "Identification of genes expressed in human CD34(+) hematopoietic
 RT stem/progenitor cells by expressed sequence tags and efficient full-
 RT length cDNA cloning.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98454325; PubMed=9781040;
 RA Rahmani Z., Maunoury C., Siddiqui A.;
 RT "Isolation of a novel human voltage-dependent anion channel gene.";
 RL Eur. J. Hum. Genet. 6:337-340 (1998).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Skin;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Rosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [4]
 RP SEQUENCE OF 1-253 FROM N.A.
 RX MEDLINE=99431679; PubMed=10501981;
 RA Decker W.K., Bowles K.R., Schatte E.C., Towbin J.A., Craigen W.J.;
 RT "Revised fine mapping of the human voltage-dependent anion channel
 RT loci by radiation hybrid analysis.";
 RL Mamm. Genome 10:1041-1042 (1999).
 RN [5]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=20293349; PubMed=10833333;
 RA Decker W.K., Craigen W.J.;
 RT "The tissue-specific, alternatively spliced single ATG exon of the
 RT type 3 voltage-dependent anion channel gene does not create a
 RT truncated protein isoform in vivo.";
 RL Mol. Genet. Metab. 70:69-74 (2000).
 CC -!- FUNCTION: Forms a channel through the mitochondrial outer membrane
 CC that allows diffusion of small hydrophilic molecules (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Outer mitochondrial membrane.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9Y277-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y277-2; Sequence=VSP_005079;
 CC -!- TISSUE SPECIFICITY: Widely expressed. Highest in testis.
 CC -!- DOMAIN: Consists mainly of membrane-spanning sided beta-sheets.
 CC -!- SIMILARITY: Belongs to the eukaryotic mitochondrial porin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AF038962; AAC39876.1; -.
 CC EMBL; U90943; AAB93872.1; -.
 CC EMBL; BC056870; AAH56870.1; -.
 CC EMBL; AF151682; AAD49610.1; -.
 CC EMBL; AF151679; AAD49610.1; JOINED.
 CC EMBL; AF151681; AAD49610.1; JOINED.
 CC IntAct; Q9Y277; -.
 CC Genew; HGNC:12674; VDAC3.
 CC GO; GO:0005741; C:mitochondrial outer membrane; TAS.
 CC GO; GO:0015482; F:voltage-dependent anion channel porin activity; TAS.
 CC GO; GO:0015853; P:anion transport; TAS.
 CC InterPro; IPR001925; Porin_Euk.
 CC Pfam; PF01459; Porin_3; 1.
 CC PRINTS; PR00185; EUKARYOTIC_PORIN.
 CC PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
 CC Alternative splicing; Mitochondrion; Outer membrane; Porin.
 KW

```
FT VARSPLIC 39 39 V -> VM (in isoform 2).
FT /FtId=VSP_005079.
SQ SEQUENCE 283 AA; 30658 MW; E03CBCEDA72A9783 CRC64;

Query Match 72.4%; Score 160; DB 1; Length 283;
Best Local Similarity 96.8%; Pred. No. 6.7e-13;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTDNTLGTISWENKLAEGKLTLDITFV 32
   |||:|||||:|||||:|||||:|||||:|||||:
Db 74 KWDNTDNTLGTISWENKLAEGKLTLDITFV 104

RESULT 4
FOR3_MOUSE STANDARD; PRT; 283 AA.
AC Q60931;
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Voltage-dependent anion-selective channel protein 3 (VDAC-3) (MVDAC3)
DE (Outer mitochondrial membrane protein porin 3).
GN Name=Vdac3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96411667; PubMed=8812436;
RA Sampson M.J., Lovell R.S., Davison D.B., Craigen W.J.;
RT "A novel mouse mitochondrial voltage-dependent anion channel gene
localizes to chromosome 8";
RL Genomics 36:192-196(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary Gland;
RX MEDLINE=22388257; PubMed=12477937; DOI=10.1073/pnas.242603899;
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
CC -!- FUNCTION: Forms a channel through the mitochondrial outer membrane
that allows diffusion of small hydrophilic molecules (By
similarity).
CC -!- SUBCELLULAR LOCATION: Outer mitochondrial membrane.
CC -!- TISSUE SPECIFICITY: Highest levels of expression detected in
testis, less but still abundant expression in heart, kidney,
brain, and skeletal muscle.
CC -!- DOMAIN: Consists mainly of membrane-spanning sided beta-sheets.
CC -!- SIMILARITY: Belongs to the eukaryotic mitochondrial porin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensee@isb-sib.ch).
CC -----
DR EMEL; U30839; AAB47776.1; -.
DR EMEL; BC004743; AA04743.1; -.
DR MGD; MG1:106922; Vdac3.
DR GO; GO:0001662; P:behavioral fear response; IMP.
DR GO; GO:0007612; P:learning; IMP.
DR GO; GO:0007270; P:nervous system synaptic transmission; IMP.
DR InterPro; IPR001925; Porin_Euk.
DR Pfam; PF01459; Porin_3; 1.
DR PRINTS; PR00185; EUKARYOTIC_PORIN.
DR PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
KW Mitochondrion; Outer membrane; Porin.
SQ SEQUENCE 283 AA; 30753 MW; A95AFD67C611228C CRC64;

Query Match 72.4%; Score 160; DB 1; Length 283;
Best Local Similarity 96.8%; Pred. No. 6.7e-13;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KWDNTDNTLGTISWENKLAEGKLTLDITFV 32
   |||:|||||:|||||:|||||:|||||:|||||:
Db 74 KWDNTDNTLGTISWENKLAEGKLTLDITFV 104

RESULT 5
Q8BNG2 PRELIMINARY; PRT; 284 AA.
AC Q8BNG2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length
DE enriched library, clone.D130008N07 product:voltage-dependent anion
DE channel 3, full insert sequence.
GN Name=Vdac3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal ganglion;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Yeth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal ganglion;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal ganglion;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal ganglion;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal ganglion;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohgami E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsushima S., Kawai J.,
RA Okazaki Y., Muranatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal ganglion;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanaigaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Kasukawa T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Tova T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK083785; BAC39019.1; -.
DR MGD; MGI:106922; Vdac3.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR GO; GO:0001562; P:behavioral fear response; IMP.
DR GO; GO:0007612; P:learning; IMP.
DR GO; GO:0007270; P:nervous system synaptic transmission; IMP.
DR InterPro; IPR001225; Porin_Euk.
DR Pfam; PF01459; Porin_3; 1.
DR PRINTS; PR00185; EUKARYOTIC PORIN.
DR PROSITE; PS00558; EUKARYOTIC PORIN; 1.
SQ SEQUENCE 284 AA; 30885 MW; 63B6AA47E4AAC4C6 CRC64;
Query Match 72.4%; Score 160; DB 2; Length 284;
Best Local Similarity 96.8%; Pred. No. 6,8e-13;
Matches 30; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KWDNTLTGTEISWENKLAEGKLTLDITFV 32
DB 75 KWDNTLTGTEISWENKLAEGKLTLDITFV 105
RESULT 6
POR3 RAT STANDARD; PRT; 283 AA.
AC Q9R1Z0; Q9BSR2; Q9J131; Q9WTU2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Voltage-dependent anion-selective channel protein 3 (VDAC-3) (rVDAC3)
GN (Outer mitochondrial membrane protein porin 3).
GN Name=Vdac3;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC TISSUE=Ascitic tumor;
RX MEDLINE=20453129; PubMed=10998069;
RA Shinohara Y., Ishida T., Hino M., Yamazaki N., Baba Y., Terada H.;
RT "Characterization of porin isoforms expressed in tumor cells.";
RL Eur. J. Biochem. 267:5067-5073(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=Sprague-Dawley;

```


OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Prostate;
RX	MEDLINE=22389257; PubMed=12477932;
RA	Strausberg R.L., Feinsold E.A., Grouse L.H., Derge J.G.,
RA	Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA	Altshul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA	Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA	Bosak S.A., McWan P.J., McKernan K.J., Malek J.A., Gurnatone P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA	Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA	Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA	Krzywinski M.I., Skalka U., Smalls D.E., Schnerch A., Schein J.E.,
RA	Jones S.J., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences."
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Prostate;
RA	Strausberg R.;
RL	Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR	EMBL; BC061780; AAH61780.1; -.
SQ	SEQUENCE 283 AA; 30784 MW; 3807231491755831 CRC64;
Query Match	71.0%; Score 157; DB 2; Length 283;
Best Local Similarity	93.5%; Pred. No. 1.7e-12;
Matches	2; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY	2 KWDNTDNLGTGEISWENKLAEGKLITDITFV 32
DB	74 KWDNTDNLGTGEISWENKLAEGKLITDITFV 104
RESULT 8	
PO3	RABIT
ID	_PO3 RABIT STANDARD; PRT; 283 AA.
AC	Q9T13;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	05-JUL-2004 (Rel. 44, Last annotation update)
DE	Voltage-dependent anion-selective channel protein 3 (VDAC-3) (Outer
DE	mitochondrial membrane protein porin 3).
GN	Name=VDAC3;
OS	Oryctolagus cuniculus (Rabbit).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Legomorpha; Leporidae; Oryctolagus.
OX	NCBI_TaxID=9986;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=New Zealand white; TISSUE=Cornea;
RA	Rae J.L.;
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC	-!- FUNCTION: Forms a channel through the mitochondrial outer membrane
CC	that allows diffusion of small hydrophilic molecules (3)
CC	similarity).
CC	-!- SUBCELLULAR LOCATION: Outer mitochondrial membrane.
CC	-!- DOMAIN: Consists mainly of eukaryotic mitochondrial porin family.
CC	-!- SIMILARITY: Belongs to the eukaryotic mitochondrial porin family.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	at the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way

```
RC TISSUE=Small intestine;
RA MEDLINE=96327607; PubMed=8672129;
RX Winterce A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine cDNA.
RL Library: analysis of 839 clones.";
RL Mamm. Genome 7:509-517(1996).
CC -!- SUBCELLULAR LOCATION: Outer mitochondrial membrane.
CC -!- DOMAIN: Consists mainly of membrane-spanning sided beta-sheets.
CC -!- SIMILARITY: Belongs to the eukaryotic mitochondrial porin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; F14590; CAA23141.1; -
DR InterPro; IPR001925; Porin_Euk.
DR Pfam; PF01459; Porin_3; 1.
DR PRINTS; PR00185; EUKARYTPORIN.
DR PROSITE; PS00558; EUKARYOTIC_PORIN; PARTIAL.
KW Mitochondrion; Outer membrane; Porin.
FT NON_TER 1
FT TER 112
FT NON_TER 112
FT TER 112
SQ SEQUENCE 112 AA; 12126 MW; 23082D5D4811593F CRC64;

Query Match 53.8%; Score 141; DB 1; Length 112;
Best Local Similarity 90.3%; Pred. No. 8.5e-11;
Matches 28; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 KWDNTLTGTSISWENKLAEGKLTLDTFV 32
DB 70 KWDNTLTGTSISWENKLAEGKLTLDTFV 100

RESULT 11
ID Q6GR11 PRELIMINARY; PRT; 283 AA.
AC Q6GR11;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]_TaxID=8355;
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
```

```
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RT Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC071123; AAH71123.1; -
DR InterPro; IPR001925; Porin_Euk.
DR Pfam; PF01459; Porin_3; 1.
DR PRINTS; PR00185; EUKARYTPORIN.
DR PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 283 AA; 30148 MW; 09EFF65765FA7918 CRC64;

Query Match 56.6%; Score 125; DB 2; Length 283;
Best Local Similarity 74.2%; Pred. No. 3.2e-08;
Matches 23; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 KWDNTLTGTSISWENKLAEGKLTLDTFV 32
DB 74 KWDNTLTGTSISWENKLAEGKLTLDTFV 104

RESULT 12
ID POR2_MELGA STANDARD; PRT; 282 AA.
AC P82013;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Voltage-dependent anion-selective channel protein 2 (VDAC-2) (Outer
DE mitochondrial membrane protein porin 2).
DN Name=VDAC2;
OS Meleagris gallopavo (Common turkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagris.
OX NCBI_TaxID=9103;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RA Hesse D., Strutz N., Kratzin H.D., Thinnies F.P., Hilschmann N.;
RL Submitted (SEP-1999) to Swiss-Prot.
CC -!- FUNCTION: Forms a channel through the cell membrane that allows
CC diffusion from small hydrophilic molecules (By similarity).
CC -!- SUBCELLULAR LOCATION: Outer mitochondrial membrane (By
CC similarity).
CC -!- DOMAIN: Consists mainly of membrane-spanning sided beta-sheets.
CC -!- SIMILARITY: Belongs to the eukaryotic mitochondrial porin family.
DR InterPro; IPR001925; Porin_Euk.
DR Pfam; PF01459; Porin_3; 1.
DR PRINTS; PR00185; EUKARYTPORIN.
DR PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
KW Acetylation; Direct protein sequencing; Mitochondrion; Outer membrane;
KW Porin.
FT MOD_RES 1 1 N-acetyllalanine.
FT BINDING 72 72 Dicyclohexylcarbodiimide.
SQ SEQUENCE 282 AA; 30066 MW; 88ECCF19ACAC004F CRC64;

Query Match 52.0%; Score 115; DB 1; Length 282;
Best Local Similarity 70.0%; Pred. No. 7e-07;
Matches 21; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 KWDNTLTGTSISWENKLAEGKLTLDTFV 31
DB 74 KWDNTLTGTSISWENKLAEGKLTLDTFV 104
```

Ds 73 KWTNTLTGTEIAIEDQIAKGLKLTFTTF 102

RESULT 13

```

POR2_XENLA
ID_POR2_XENLA STANDARD; PRT; 282 AA.
AC P81004;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Voltage-dependent anion-selective channel protein 2 (Outer
DE mitochondrial membrane protein porin).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
RN [1]
RP SEQUENCE.
RC TISSUE=Skeletal muscle;
RA Reymann S., Kratzin H.D., Hesse D., Hesse J.-O., Klebert S.,
RA Klafard Z., Zimmermann B., Spengler B., Metzger S., Thinnies F.P.,
RA Hilschmann N.;
RA Submitted (JUL-1997) to Swiss-Prot.
RN [2]
RP SEQUENCE OF 84-90 AND 274-282, AND TISSUE SPECIFICITY.
RC TISSUE=Oocyte;
RX MEDLINE=20150664; PubMed=10687956;
RA Steinacker P., Awai L.A., Becker S., Cole T., Reymann S., Hesse D.,
RA Kratzin H.D., Morris-Wortmann C., Schwarzer C., Thinnies F.P.,
RA Hilschmann N.;
RT "The plasma membrane of Xenopus laevis oocytes contains voltage-
RT dependent anion-selective porin channels.";
RL Int. J. Biochem. Cell Biol. 32:225-234 (2000).
CC -1- FUNCTION: Forms a channel through the mitochondrial outer membrane
CC that allows diffusion of small hydrophilic molecules. The channel
CC adopts an open conformation at low or zero membrane potential and
CC has a closed conformation at potentials above 30-40 mV. The open state
CC has a weak anion selectivity whereas the closed state is cation-
CC selective (By similarity).
CC -1- SUBCELLULAR LOCATION: Outer mitochondrial membrane (Potential).
CC -1- TISSUE SPECIFICITY: Expressed in skeletal muscle and oocytes.
CC -1- DOMAIN: Consists mainly of membrane-spanning sided beta-sheets.
CC -1- SIMILARITY: Belongs to the eukaryotic mitochondrial porin family.
DR InterPro: IPR001925; Porin_3; 1.
DR Pfam: PF01459; Porin_3; 1.
DR PRINTS: PR00185; EUKARYOTIC_PORIN.
DR PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
KW Acetylation; Direct protein sequencing; Mitochondrion; Outer membrane;
KW Porin.
FT MOD RES 1 1 N-acetylalanine.
SQ SEQUENCE 282 AA; 30070 MW; B0309215D81FF313 CRC64;

Query Match 52.0%; Score 115; DB 1; Length 282;
Best Local Similarity 70.0%; Pred. No. 7e-07;
Matches 21; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 KWTNTLTGTEISWENKLAEGKLTLDITF 31
|||:|||||:|||||:|||||:|||||:
Db 73 KWTNTLTGTEIAIEDQIAKGLKLTFTTF 102
|||:|||||:|||||:|||||:|||||:

RESULT 14
Q8AWD0
ID Q8AWD0 PRELIMINARY; PRT; 283 AA.
AC Q8AWD0;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE Hypothetical protein zgc:55795 (voltage-dependent anion channel
DE 2).
DN Name=zgc:55795;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole, and Whole body;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lomuellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RA Strausberg R.;
RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney marrow;
RA Song H.D., Wu X.Y., Sun X.J., Zhou Y., Liu T.X., Deng M., Zhang G.W.,
RA Sheng Y., Chen Y., Ruan Z., Jiang C.L., Fan H.Y., Zou L.I.,
RA Xanki J.P., Look A.T., Chen Z.;
RA Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RA Strausberg R.;
RA Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042329; AAH42329.1; -.
DR EMBL; AY394955; AAQ94582.1; -.
DR EMBL; BC062525; AAH62525.1; -.
DR GO; GO:0005741; C:mitochondrial outer membrane; IEA.
DR GO; GO:0008308; F:voltage-dependent ion-selective channel act. .; IEA.
DR GO; GO:0006820; P:anion transport; IEA.
DR InterPro: IPR001925; Porin_Euk.
DR Pfam: PF01459; Porin_3; 1.
DR PRINTS: PR00185; EUKARYOTIC_PORIN.
KW Hypothetical protein.
SQ SEQUENCE 283 AA; 30284 MW; 935494C3B48DFB41 CRC64;

Query Match 52.0%; Score 115; DB 2; Length 283;
Best Local Similarity 70.0%; Pred. No. 7e-07;
Matches 21; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 KWTNTLTGTEISWENKLAEGKLTLDITF 31
|||:|||||:|||||:|||||:|||||:
Db 74 KWTNTLTGTEINTDQIAKGLKLTFTTF 103
|||:|||||:|||||:|||||:|||||:

RESULT 15
Q919D1
ID Q919D1 PRELIMINARY; PRT; 283 AA.
AC Q919D1;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
```

DE Voltage-dependent anion channel.
 GN Name=VDAC2;
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=breed White Leghorn; TISSUE=Lens fiber;
 RA Rae J.L.; (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF268470; AAF73513.1;
 DR GO; GO:0005741; C:mitochondrial outer membrane; IEA.
 DR GO; GO:0008308; F:voltage-dependent ion-selective channel act. . .; IEA.
 DR GO; GO:0006820; P:anion transport; IEA.
 DR InterPro; IPR001925; Porin_Euk.
 DR Pfam; PF01459; Porin_3; 1.
 DR PRINTS; PRO0185; EUKARYOTIC_PORIN.
 DR PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
 SQ SEQUENCE 283 AA; 30198 MW; 4DB5DC020A632902 CRC64;

Query Match 52.0%; Score 115; DB 2; Length 283;
 Best Local Similarity 70.0%; Pred. No. 7e-07;
 Matches 21; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 KWDTDNTLGTETISWENKLAEGKLTLDITF 31
 |||:|||||:|:|:|:|:|:|
 Db 74 KWDTDNTLGTETIAIEDQIAKGLKLTDTTF 103

Search completed: November 10, 2004, 12:27:05
 Job time : 44.7654 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 9.23302 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-13
Perfect score: 156
Sequence: 1 RGAVFSQDKVQVQATKVLNRNADNFYINDR 31

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PGTUS COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84	53.8	152	4	US-09-270-767-57836
2	84	53.8	592	4	US-09-270-767-42535
3	53	34.0	872	1	US-08-451-715A-8
4	52	33.3	191	4	US-09-248-796A-24047
5	50	32.1	195	3	US-09-117-257-46
6	50	32.1	195	3	US-09-489-352-46
7	48	30.8	340	4	US-09-543-681A-7558
8	47	30.1	160	3	US-09-117-257-38
9	47	30.1	160	3	US-09-489-352-38
10	47	30.1	194	3	US-09-117-257-17
11	47	30.1	194	3	US-08-945-476-17
12	47	30.1	194	3	US-09-489-352-17
13	47	30.1	195	3	US-09-117-257-34
14	47	30.1	195	3	US-09-489-352-34
15	47	30.1	398	4	US-09-248-796A-15060
16	47	30.1	418	4	US-09-248-796A-20578
17	47	30.1	855	4	US-09-248-796A-15487
18	46.5	29.8	186	4	US-09-248-796A-15793
19	46	29.5	327	4	US-09-595-857B-27
20	46	29.5	358	4	US-09-861-451A-50
21	46	29.5	395	3	US-09-088-216-4
22	46	29.5	427	4	US-09-252-991A-19310
23	46	29.5	469	3	US-08-378-313-33
24	46	29.5	1457	3	US-08-665-259-27
25	46	29.5	1457	3	US-08-762-500-27
26	46	29.5	1472	4	US-09-032-438C-119
27	46	29.5	2310	4	US-09-874-923-120

28	45.5	29.2	227	4	US-09-328-352-8175	Sequence 8175, Ap
29	45	28.8	97	4	US-09-328-352-6261	Sequence 6261, Ap
30	45	28.8	257	4	US-09-809-665A-61	Sequence 61, Appl
31	45	28.8	266	3	US-09-134-001C-4921	Sequence 4921, Ap
32	45	28.8	401	4	US-09-540-236-3566	Sequence 3566, Ap
33	45	28.8	467	4	US-09-363-243-3	Sequence 3, Appl
34	45	28.8	1684	3	US-08-665-259-25	Sequence 25, Appl
35	45	28.8	1684	3	US-08-762-500-25	Sequence 25, Appl
36	45	28.8	1704	3	US-08-762-500-75	Sequence 75, Appl
37	44.5	28.5	271	4	US-09-107-532A-3746	Sequence 3746, Ap
38	44.5	28.5	566	4	US-09-538-092-807	Sequence 807, App
39	44	28.2	118	3	US-08-858-207A-395	Sequence 395, App
40	44	28.2	127	1	US-08-259-696B-2	Sequence 2, Appl
41	44	28.2	127	2	US-08-902-513-2	Sequence 2, Appl
42	44	28.2	128	1	US-08-259-696B-3	Sequence 3, Appl
43	44	28.2	128	2	US-08-902-513-3	Sequence 3, Appl
44	44	28.2	140	4	US-09-134-000C-4822	Sequence 4822, Ap
45	44	28.2	144	1	US-08-259-696B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-270-767-57836
; Sequence 57836, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 57836
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-57836

Query Match 53.8%; Score 84; DB 4; Length 152;
Best Local Similarity 56.7%; Pred. No. 2.5e-05;
Matches 17; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 GAVFSQDKVQVQATKVLNRNADNFYINDR 31
DB 68 GAVFGQDEDFVKCALQEFPMAGNFYINDK 97

RESULT 2
US-09-270-767-42535
; Sequence 42535, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 42535
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-42535

Query Match 53.8%; Score 84; DB 4; Length 592;
Best Local Similarity 56.7%; Pred. No. 0.00014;
Matches 17; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 GAVFSQDKVQVQATKVLNRNADNFYINDR 31

; ORGANISM: Borrelia burgdorferi
US-09-489-352-46

Query Match 32.1%; Score 50; DB 3; Length 195;
Best Local Similarity 47.1%; Pred. No. 6.5;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 9 KDVQVQATKVLNRNADN 25
DB 43 KDIIDEINKIKKDAADN 59

RESULT 7
US-09-543-681A-7558
; Sequence 7558, Application US/09543681A
; Patent No. 6805709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7558
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7558

Query Match 30.8%; Score 48; DB 4; Length 340;
Best Local Similarity 33.3%; Pred. No. 26;
Matches 10; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 2 GAVFSQKXVQVQATKVLNRNADNFYINDR 31
DB 70 GAVINCAKQVQVQALNARKEQMESALINER 99

RESULT 8
US-09-117-257-38
; Sequence 38, Application US/09117257
; Patent No. 6214355
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Guo, Betty
; APPLICANT: Hanson, Mark
; TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 4210.000500
; CURRENT APPLICATION NUMBER: US/09/117,257
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: PCT/US96/17081
; EARLIER FILING DATE: 1996-10-22
; EARLIER APPLICATION NUMBER: 08/589,711
; EARLIER FILING DATE: 1996-01-22
; EARLIER APPLICATION NUMBER: 08/427,023
; EARLIER FILING DATE: 1995-04-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-09-117-257-38

Query Match 30.1%; Score 47; DB 3; Length 160;
Best Local Similarity 47.1%; Pred. No. 15;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 9 KDVQVQATKVLNRNADN 25

Db 43 KDIIDEINKIKKDAADN 59

RESULT 9
US-09-489-352-38
; Sequence 38, Application US/09489352
; Patent No. 6312907
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Guo, Betty
; APPLICANT: Hanson, Mark
; TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 4210.000500
; CURRENT APPLICATION NUMBER: US/09/489,352
; CURRENT FILING DATE: 2000-01-21
; EARLIER APPLICATION NUMBER: PCT/US96/17081
; EARLIER FILING DATE: 1996-10-22
; EARLIER APPLICATION NUMBER: 08/589,711
; EARLIER FILING DATE: 1996-01-22
; EARLIER APPLICATION NUMBER: 08/427,023
; EARLIER FILING DATE: 1995-04-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-09-489-352-38

Query Match 30.1%; Score 47; DB 3; Length 160;
Best Local Similarity 47.1%; Pred. No. 15;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 9 KDVQVQATKVLNRNADN 25
DB 43 KDIIDEINKIKKDAADN 59

RESULT 10
US-09-117-257-17
; Sequence 17, Application US/09117257
; Patent No. 6214355
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Guo, Betty
; APPLICANT: Hanson, Mark
; TITLE OF INVENTION: DbpA AND DbpB COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 4210.000500
; CURRENT APPLICATION NUMBER: US/09/117,257
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: PCT/US96/17081
; EARLIER FILING DATE: 1996-10-22
; EARLIER APPLICATION NUMBER: 08/589,711
; EARLIER FILING DATE: 1996-01-22
; EARLIER APPLICATION NUMBER: 08/427,023
; EARLIER FILING DATE: 1995-04-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-09-117-257-17

Query Match 30.1%; Score 47; DB 3; Length 194;
Best Local Similarity 47.1%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 9 KDVQVQATKVLNRNADN 25
DB 43 KDIIDEINKIKKDAADN 59

```

RESULT 13
US-09-117-257-34
; Sequence 34, Application US/09117257
; Patent No. 6214355
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Guo, Betty
; APPLICANT: Hanson, Mark
; TITLE OF INVENTION: DPGA AND DbpB COMPOSITIONS AND METHODS OF USE
; FILE REFERENCE: 4210.000900
; CURRENT APPLICATION NUMBER: US/09/117,257
; CURRENT FILING DATE: 1998-07-22
; EARLIER APPLICATION NUMBER: PCT/US96/17081
; EARLIER FILING DATE: 1996-10-22
; EARLIER APPLICATION NUMBER: 08/589,711
; EARLIER FILING DATE: 1996-01-22
; EARLIER APPLICATION NUMBER: 08/427,023
; EARLIER FILING DATE: 1995-04-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 34
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-09-117-257-34

```

Query Match	30.1%	Score 47;	DB 3;	Length 195;
Best Local Similarity	47.1%	Pred. NO. 19;		
Matches	8;	Conservative	4;	Mismatches 5;
Indels				0;
Gaps				0;

RESULT 14
US-09-489-352-34
 ; Sequence 34, Application US/09489352
 ; Patent No. 6312907
 ; GENERAL INFORMATION:
 ; APPLICANT: Hook, Magnus
 ; APPLICANT: Guo, Betty
 ; APPLICANT: Hanson, Mark
 ; TITLE OF INVENTION: DbPa AND DbPb COMPOSITIONS AND METHODS OF USE
 ; FILE REFERENCE: 4210.000500
 ; CURRENT APPLICATION NUMBER: US/09/489,352
 ; CURRENT FILING DATE: 2000-01-21
 ; EARLIER APPLICATION NUMBER: PCT/US96/17081
 ; EARLIER FILING DATE: 1996-10-22
 ; EARLIER APPLICATION NUMBER: 08/589,711
 ; EARLIER FILING DATE: 1996-01-22
 ; EARLIER APPLICATION NUMBER: 08/427,023
 ; EARLIER FILING DATE: 1995-04-24
 ; NUMBER OF SEQ ID NOS: 66
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 34
 ; LENGTH: 195
 ; TYPE: PRT
 ; ORGANISM: Borrelia burgdorferi
US-09-489-352-34

```

Query Match      30.1%; Score 47; DB 3; Length 195;
Best Local Similarity 47.1%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Qy      9 KDVTVEATKVLRNAADN 25
      ||| :| :|||
Db      43 KDITDEINKKDAADN 59

RESULT 15

```


US-09-248-796A-15060
; Sequence 15060, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15060
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-15060

Query Match 30.1%; Score 47; DB 4; Length 398;
Best Local Similarity 35.3%; Pred. No. 46;
Matches 12; Conservative 7; Mismatches 11; Indels 4; Gaps 1;

QY 2 GAVFSQDKDVVQEQATKVLNRNADNFY----INDR 31
Db 310 GNLTKDKKXFLMLKILNSRDNYLSLEINDK 343

Search completed: November 10, 2004, 12:32:27
Job time : 10.233 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 15:53:52 ; Search time 29.8997 Seconds
(without alignments)
366.225 Million cell updates/sec

Title: US-10-092-750-13

Perfect score: 156

Sequence: 1 RGAVSQDKDVQVEATKVLNRNADNFYINDR 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- Published Applications AA:*
- 1: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pdb:*
 - 2: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pdb:*
 - 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pdb:*
 - 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pdb:*
 - 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pdb:*
 - 6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pdb:*
 - 7: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pdb:*
 - 8: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pdb:*
 - 9: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pdb:*
 - 10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pdb:*
 - 11: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pdb:*
 - 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pdb:*
 - 13: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pdb:*
 - 14: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pdb:*
 - 15: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pdb:*
 - 16: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pdb:*
 - 17: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pdb:*
 - 18: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pdb:*
 - 19: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pdb:*
 - 20: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pdb:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156	100.0	31	14	US-10-092-750-13
2	141	90.4	308	14	Sequence 13, Appl
3	141	90.4	563	16	Sequence 384, App
4	96	61.5	548	14	Sequence 1800, Ap
5	90	57.7	553	14	Sequence 2507, Ap
6	89	57.1	622	14	Sequence 22859, A
7	88	56.4	544	14	Sequence 13433, A
8	80	51.3	552	14	Sequence 19524, A
9	78	50.0	402	14	Sequence 3745, Ap
10	77	49.4	211	14	Sequence 8341, Ap
11	77	49.4	575	14	Sequence 18, Appl
12	76	48.7	542	14	Sequence 22032, A
13	67	42.9	543	14	Sequence 10453, A
					Sequence 10260, A

14	53.5	34.3	537	14	US-10-369-493-3433	Sequence 3433, Ap
15	53	34.0	241	15	US-10-282-122A-52577	Sequence 52577, A
16	53	34.0	874	10	US-09-882-227-262	Sequence 262, App
17	53	34.0	874	15	US-10-282-122A-58891	Sequence 58891, A
18	53	34.0	1134	15	US-10-282-122A-53179	Sequence 53179, A
19	52	33.3	469	14	US-10-369-493-11085	Sequence 11085, A
20	51	32.7	2231	14	US-10-369-493-1830	Sequence 1830, Ap
21	50.5	32.4	687	14	US-10-369-493-2331	Sequence 2331, Ap
22	50.5	32.4	718	15	US-10-282-122A-76800	Sequence 76800, A
23	50	32.1	241	9	US-09-815-242-13907	Sequence 13907, A
24	50	32.1	241	15	US-10-282-122A-72873	Sequence 72873, A
25	50	32.1	241	15	US-10-282-122A-75272	Sequence 75272, A
26	50	32.1	241	15	US-10-282-122A-76144	Sequence 76144, A
27	49.5	31.7	252	17	US-10-739-930-10399	Sequence 10399, A
28	49	31.4	241	15	US-10-282-122A-55990	Sequence 55990, A
29	49	31.4	504	16	US-10-767-701-43996	Sequence 43996, A
30	49	31.4	930	16	US-10-767-701-43996	Sequence 43996, A
31	49	31.4	1138	16	US-10-437-963-124691	Sequence 124691, A
32	49	31.4	1710	16	US-10-437-963-153401	Sequence 153401, A
33	48.5	31.1	116	14	US-10-369-493-19557	Sequence 19557, A
34	48.5	31.1	2735	16	US-10-437-963-182452	Sequence 182452, A
35	48	30.8	186	16	US-10-767-701-42655	Sequence 42655, A
36	48	30.8	189	17	US-10-425-115-236019	Sequence 236019, A
37	48	30.8	224	17	US-10-425-115-33113	Sequence 33113, A
38	48	30.8	226	9	US-09-815-242-11016	Sequence 11016, A
39	48	30.8	226	15	US-10-282-122A-58161	Sequence 58161, A
40	48	30.8	227	15	US-10-424-599-145066	Sequence 145066, A
41	48	30.8	259	16	US-10-767-701-35954	Sequence 35954, A
42	48	30.8	306	15	US-10-335-977-7083	Sequence 7083, Ap
43	48	30.8	327	15	US-10-282-122A-68588	Sequence 68588, A
44	48	30.8	341	15	US-10-335-977-7084	Sequence 7084, Ap
45	48	30.8	385	15	US-10-425-114-39635	Sequence 39635, A

ALIGNMENTS

RESULT 1
US-10-092-750-13
; Sequence 13, Application US/10092750
; Publication No. US2003032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-XL
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-13

Query Match 100.0%; Score 156; DB 14; Length 31;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGAVSQDKDVQVEATKVLNRNADNFYINDR 31
DB 1 RGAVSQDKDVQVEATKVLNRNADNFYINDR 31

RESULT 2
US-10-043-487-384
; Sequence 384, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:

```
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, LÉGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptid
; FILE OF INVENTION: mammalian polypeptides
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 384
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Shigella Flexneri
US-10-043-487-384

Query Match          90.4%; Score 141; DB 14; Length 308;
Best Local Similarity 93.3%; Pred. No. 2.7e-12;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAVFSQDKDVQVEATKVLNRNADNFYINDR 31
DB 225 GAVFSQDKDVQVEATKVLNRNAGNFYINDK 254

RESULT 3
US-10-408-765A-1800
; Sequence 1800, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Faby, Boin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088,465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1800
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1800

Query Match          90.4%; Score 141; DB 16; Length 563;
Best Local Similarity 93.3%; Pred. No. 5.7e-12;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAVFSQDKDVQVEATKVLNRNADNFYINDR 31
DB 480 GAVFSQDKDVQVEATKVLNRNAGNFYINDK 509

RESULT 4
US-10-369-493-2507
; Sequence 2507, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; LENGTH: 622

US-10-092-750-13.rapb

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2507

Query Match          61.5%; Score 96; DB 14; Length 548;
Best Local Similarity 60.0%; Pred. No. 2.3e-05;
Matches 18; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 GAVFSQDKDVQVEATKVLNRNADNFYINDR 31
DB 465 GSIFAQDRVVVKLTDLRLNRNAGNFYINDK 494

RESULT 5
US-10-369-493-22859
; Sequence 22859, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22859
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Aeropyrum pernix
US-10-369-493-22859

Query Match          57.7%; Score 90; DB 14; Length 553;
Best Local Similarity 53.3%; Pred. No. 0.00018;
Matches 16; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 2 GAVFSQDKDVQVEATKVLNRNADNFYINDR 31
DB 468 GSVFAKDREAILKAEKALRYAAGNFYINDK 497

RESULT 6
US-10-369-493-13433
; Sequence 13433, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; LENGTH: 622
```

```
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(622)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-13433

Query Match      57.1%; Score 89; DB 14; Length 622;
Best Local Similarity 60.7%; Pred. No. 0.00029;
Matches 17; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY      2 GAVFSQDKDVQVQKATKVLNRNADNFYIN 29
Db      538 GSVFAQDREALAVDVLNRNAGNFYIN 565

RESULT 7
US-10-369-493-19524
; Sequence 19524, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19524
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-10-369-493-19524

Query Match      56.4%; Score 88; DB 14; Length 544;
Best Local Similarity 53.3%; Pred. No. 0.00034;
Matches 16; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY      2 GAVFSQDKDVQVQKATKVLNRNADNFYINR 31
Db      459 GAVFARDRKAIIDTALRELRNAGNFYINDK 488

RESULT 8
US-10-369-493-3745
; Sequence 3745, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3745
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
```

```
; NAME/KEY: unsure
; LOCATION: (1)..(552)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-3745

Query Match      51.3%; Score 80; DB 14; Length 552;
Best Local Similarity 50.0%; Pred. No. 0.0053;
Matches 14; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY      2 GAVFSQDKDVQVQKATKVLNRNADNFYIN 29
Db      470 GSIFASDRSVIRFAEKLKRLNSAGNFYIN 497

RESULT 9
US-10-369-493-8341
; Sequence 8341, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8341
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-10-369-493-8341

Query Match      50.0%; Score 78; DB 14; Length 402;
Best Local Similarity 50.0%; Pred. No. 0.0071;
Matches 15; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY      2 GAVFSQDKDVQVQKATKVLNRNADNFYINR 31
Db      321 GAIIATDRAAILAATRLRFAAGNFYINDK 350

RESULT 10
US-10-230-331-18
; Sequence 18, Application US/10230331
; Publication No. US20030157513A1
; GENERAL INFORMATION:
; APPLICANT: RAJASEKHARAN, Ram
; TITLE OF INVENTION: A NOVEL TRIACYLGLYCEROL BIOSYNTHESIS IN THE CYTOSOL OF EUKARYOTES
; FILE REFERENCE: 110522
; CURRENT APPLICATION NUMBER: US/10/230,331
; CURRENT FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: US 60/315,757
; PRIOR FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Diacylglycerol Acyltransferase Protein Sequence
US-10-230-331-18

Query Match      49.4%; Score 77; DB 14; Length 211;
Best Local Similarity 46.7%; Pred. No. 0.0045;
Matches 14; Conservative 7; Mismatches 9; Indels 0; Gaps 0;
```

QY 2 GAVFSQDKVQVEATKVLNRNADNFYINDR 31
 Db 144 GCIFSDRATVTKAGALLRHAAGNYINDK 173

RESULT 11
 US-10-369-493-22032
 ; Sequence 22032, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CAO, YONGWEI
 ; APPLICANT: HINKLE, GREGORY J.
 ; APPLICANT: SLATER, STEVEN C.
 ; APPLICANT: GOLDMAN, BARRY S.
 ; APPLICANT: CHEN, XIANFENG
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 22032
 ; LENGTH: 575
 ; TYPE: PRT
 ; ORGANISM: Saccharomyces cerevisiae
 US-10-369-493-22032

Query Match 49.4%; Score 77; DB 14; Length 575;
 Best Local Similarity 43.3%; Pred. No. 0.015; 7; Indels 0; Gaps 0;
 Matches 13; Conservative 10; Mismatches 7;

QY 2 GAVFSQDKVQVEATKVLNRNADNFYINDR 31
 Db 491 GAIFAKDKRAIEYADEKLFSGAGNYINDK 520

RESULT 12
 US-10-369-493-10453
 ; Sequence 10453, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CAO, YONGWEI
 ; APPLICANT: HINKLE, GREGORY J.
 ; APPLICANT: SLATER, STEVEN C.
 ; APPLICANT: GOLDMAN, BARRY S.
 ; APPLICANT: CHEN, XIANFENG
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 10453
 ; LENGTH: 542
 ; TYPE: PRT
 ; ORGANISM: Cytophaga hutchinsonii
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(542)
 ; OTHER INFORMATION: unsure at all Xaa locations
 US-10-369-493-10453

Query Match 48.7%; Score 76; DB 14; Length 542;
 Best Local Similarity 43.3%; Pred. No. 0.02;
 Matches 13; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 2 GAVFSQDKVQVEATKVLNRNADNFYINDR 31
 Db 460 GSIFADSAIEAEKALYYSAGNYINDK 489

RESULT 13
 US-10-156-761-10260
 ; Sequence 10260, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 10260
 ; LENGTH: 543
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-10260

Query Match 42.9%; Score 67; DB 14; Length 543;
 Best Local Similarity 50.0%; Pred. No. 0.42;
 Matches 15; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 2 GAVFSQDKVQVEATKVLNRNADNFYINDR 31
 Db 460 GAVSNDRAAAAYTMDKLRVYAGNYINDK 489

RESULT 14
 US-10-369-493-3433
 ; Sequence 3433, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CAO, YONGWEI
 ; APPLICANT: HINKLE, GREGORY J.
 ; APPLICANT: SLATER, STEVEN C.
 ; APPLICANT: GOLDMAN, BARRY S.
 ; APPLICANT: CHEN, XIANFENG
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 3433
 ; LENGTH: 537
 ; TYPE: PRT
 ; ORGANISM: Neurospora crassa
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(537)
 ; OTHER INFORMATION: unsure at all Xaa locations
 US-10-369-493-3433

Query Match 34.3%; Score 53.5; DB 14; Length 537;
 Best Local Similarity 43.3%; Pred. No. 40;
 Matches 13; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

QY 1 RGAVFSQDKVQVEATKVLNRNADNFYIND 30
 Db 335 RVTVPFQEDVTRTATVLRGATQN-HLDD 363

```

RESULT 15
US-10-282-122A-52577
; Sequence 52577, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Orlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52577:
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52577

Query Match 34.0%; Score 53; DB 15; Length 241;
Best Local Similarity 40.0%; Pred. No. 18;
Matches 10; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 4 VFSQDKDVQEAATKVLNADNFYI 28
||| : : : : :
Db 70 VFSDDLETMEKEAEVLGNLGENVYI 94

Search completed: November 10, 2004, 16:35:49
Job time : 30.8997 secs

```

This Page Blank (uspio)


```

C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: A64664
R;Tomb, J.F.; White, O.; Karlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.N. A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori. A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: A64664
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-874 <TOM>
A;Cross-references: UNIPROT:P56000; GB:AE000621; GB:AE000511; NID:g2314301; PIDN:AAD08195
C;Superfamily: valine-tRNA ligase
C;Keywords: ligase

Query Match 34.0%; Score 53; DB 2; Length 874;
Best Local Similarity 50.0%; Pred. No. 27;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 9 KDVQVEATKVLNRNADNFYND 30
| : ||| ||| ||| :
Db 619 KSLNSATKEARNALDNYRND 640

RESULT 10
T21712
Hypothetical protein F33H2.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T21712
R;Cottage, A. submitted to the EMBL Data Library, November 1996
A;Reference number: Z19463
A;Accession: T21712
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-214 <MIL>
A;Cross-references: UNIPROT:O62218; EMBL:Z81526; PIDN:CAB04263.1; GSPDB:GN00019; CESP:F33
A;Experimental source: clone F33H2
C;Genetics:
A;Gene: CESP:F33H2.5
A;Map position: 1
A;Introns: 14/2; 227/2; 337/3; 544/2; 626/3; 661/1; 776/1; 964/3; 1153/2; 1720/2; 1875/3;
C;Superfamily: DNA-directed DNA polymerase II

Query Match 34.0%; Score 53; DB 2; Length 2144;
Best Local Similarity 41.7%; Pred. No. 74;
Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 7 QDKDVQVEATKVLNRNADNFYND 30
| : ||| ||| :
Db 54 QPSLVDQTKVLIISVVDYFISD 77

RESULT 11
H82492
conserved hypothetical protein VCA0167 [imported] - Vibrio cholerae (strain N16961 serog;
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: H82492
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P. l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae. A;Reference number: A82035; MUID:20405833; PMID:10952301
A;Accession: H82492
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <HEI>

```

```

A;Cross-references: UNIPROT:Q9RM29; GB:AB004357; GB:AB003853; NID:g9657547; PIDN:AAF9608
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0167
A;Map position: 2

Query Match      33.7%; Score 52.5; DB 2; Length 262;
Best Local Similarity 30.2%; Pred. No. 8.1;
Matches 13; Conservative 5; Mismatches 17; Gaps 1;

Qy      1  RGAVFSQD-----KDVQVQATKVLNRNADNF 26
Db      174  QGVVFNQDPNNRVFSTGLNAIKPDMIEATKQAREVAQKF 216

RESULT 12
A11752
endopeptidase [bacteriophage bTL285] homolog lin2566 [imported] - Listeria innocua (stra
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: A11752
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournan, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: A11752
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-713 <GLA>
A;Cross-references: UNIPROT:Q928G9; GB:AL592022; PIDN:CAC97793.1; PID:gl6415088; GSPDB:G
A;Experimental source: strain Clp11262
C;Genetics:
A;Gene: lin2566

Query Match      33.3%; Score 52; DB 2; Length 713;
Best Local Similarity 47.8%; Pred. No. 30;
Matches 11; Conservative 6; Mismatches 4; Indels 2; Gaps 1;

Qy      5  FSQDKVQVQATKVLNRNADNFY 27
Db      256  YSSDENVVDRIT--LRNASKNYF 276

RESULT 13
T28652
erythrocyte binding protein - Plasmodium knowlesi
C;Species: plasmodium knowlesi
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T28652
R;Adams, J.H.; Sim, B.K.; Dolan, S.A.; Fang, X.; Kaslow, D.C.; Miller, L.H.
Proc. Natl. Acad. Sci. U.S.A. 89, 7085-7089, 1992
A;Title: A family of erythrocyte binding proteins of malaria parasites.
A;Reference number: Z20495; MUID:92357776; PMID:1496004
A;Accession: T28652
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1153 <ADA>
A;Cross-references: UNIPROT:P50493; EMBL:M90694; NID:gl60290; PID:gl60291; PIDN:AAA29603
C;Genetics:
A;Introns: 20/3; 1087/3; 1114/2

Query Match      32.7%; Score 51; DB 2; Length 1153;
Best Local Similarity 42.9%; Pred. No. 71;
Matches 9; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy      11  VVOEATKVLNRNADNFYINDR 31
Db      105  ILKESTYEAQNVAADNYYIDDK 125

```

```

RESULT 14
S53416
SEN1 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein I9576.1; protein YLR430W
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S53416; A44387; S41985
R;Favell, A.
submitted to the EMBL Data Library, February 1995
A;Description: The sequence of S. cerevisiae cosmid 9576.
A;Reference number: S53409
A;Accession: S53416
A;Molecule type: DNA
A;Residues: 1-2231 <FAV>
A;Cross-references: UNIPROT:Q00416; EMBL:U20939; NID:g664871; PID:g664872; MIPS:YLR430W
A;Experimental source: strain S288C (AB972)
R;deMarini, D.J.; Wiley, M.; Ursic, D.; Webb, F.; Culbertson, M.R.
Mol. Cell. Biol. 12, 2154-2164, 1992
A;Title: SEN1, a positive effector of tRNA-splicing endonuclease in Saccharomyces cerevis
A;Reference number: A44387; MUID:92236590; PMID:1569945
A;Accession: A44387
A;Molecule type: DNA
A;Residues: 'MHS', 130, 'FCEREVQ', 131-2231 <DEM>
A;Cross-references: GB:M74589; NID:gl72573; PIDN:AAB63976.1; PID:gl72574
C;Genetics:
A;Gene: SGD:SEN1
A;Cross-references: SGD:S0004422; MIPS:YLR430W
A;Map position: 12R
C;Function:
A;Description: may be component of nuclear splicing complex
C;Keywords: nucleotide binding; nucleus; P-loop
F1357-1364/Region: nucleotide-binding motif A (P-loop)

Query Match      32.7%; Score 51; DB 2; Length 2231;
Best Local Similarity 47.6%; Pred. No. 1.5e+02;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy      4  VFSQDKVQVQATKVLNRNAD 24
Db      673  IFSSDKHLYQAATNLYNTFD 693

RESULT 15
T40258
Probable t-complex protein 1, theta subunit - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T40258
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z21868
A;Accession: T40258
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-546 <LYN>
A;Cross-references: UNIPROT:P78921; EMBL:AL031854; PIDN:CAA21275.1; GSPDB:GN00067; SPDB:
A;Experimental source: strain 972h-; cosmid c337
C;Genetics:
A;Gene: SPDB:SPBC337.05C
A;Map position: 2
A;Introns: 41/3; 55/1; 95/3
C;Superfamily: molecular chaperone t-complex-type

Query Match      32.4%; Score 50.5; DB 2; Length 546;
Best Local Similarity 43.3%; Pred. No. 36;
Matches 13; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Qy      1  RGAVFSQDKVQVQATKVLNRNADNFYIND 30
Db      359  RVTVFRQVEDITRTATIVLRGATKT-YLDD 387

```

us-10-092-750-13.rpr

Fri Nov 12 14:54:49 2004

Search completed: November 10, 2004, 12:29:16
Job time : 7.26698 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01 ; Search time 33.9182 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: us-10-092-750-13

Perfect score: 156

Sequence: 1 RGAFSQKDVQVQKATKURNADNFYNDR 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Watch 0%

Maximum Watch 100%

Listing first 45 summaries

Database : Uniprot_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	141	90.4	563	1 PUT2_HUMAN	P30038 homo sapien
2	141	90.4	563	2 Q81Z38	Q81Z38 homo sapien
3	141	90.4	563	2 Q961F0	Q961F0 homo sapien
4	124	79.5	1465	2 Q7TP27	Q7TP27 rattus norv
5	123	78.8	381	2 Q8R0N1	Q8R0N1 mus musculu
6	123	78.8	549	2 Q8RLS2	Q8RLS2 mus musculu
7	123	78.8	562	2 Q8BXN3	Q8BXN3 mus musculu
8	123	78.8	562	2 Q8CHT0	Q8CHT0 mus musculu
9	111	71.2	556	2 Q7SY23	Q7SY23 brachydanio
10	96	61.5	548	1 PUT2_SCHPO	Q74766 schizosacch
11	90	57.7	553	2 Q9YD2	Q9YD2 aeropyrum p
12	89	57.1	553	2 Q16448	Q16448 caenorhabdi
13	89	57.1	572	1 PUT5_EMENI	Q9P810 emeritella
14	88	56.4	560	2 Q9NKR5	Q9NKR5 leishmania
15	87	55.8	573	2 Q7SEZ9	Q7SEZ9 ashbya gos
16	87	55.8	573	2 AAS50295	AAS50295 ashbya go
17	85	54.5	572	2 Q6CF74	Q6CF74 yarrowia li
18	84	53.8	546	2 Q7V818	Q7V818 bordetella
19	84	53.8	574	2 Q6FW05	Q6FW05 candida gla
20	84	53.8	574	2 Q9VNX4	Q9VNX4 drosophila
21	84	53.8	574	2 Q8T3P0	Q8T3P0 drosophila
22	82	52.6	543	2 Q7MW36	Q7MW36 porphyromon
23	82	52.6	586	2 Q6BH88	Q6BH88 debaryomyce
24	80	51.3	601	2 Q7SDT4	Q7SDT4 neurospora
25	77	49.4	575	1 PUT2_YEAST	P07275 saccharomyc
26	76	48.7	576	2 Q7Z458	Q7Z458 kluyveromyc
27	72	46.2	543	2 C50443	C50443 mycobacteri
28	72	46.2	543	2 Q7U0F8	Q7U0F8 mycobacteri
29	72	46.2	607	2 Q7QA89	Q7QA89 anopheles g
30	70	44.9	546	2 Q73WR7	Q73WR7 mycobacteri
31	70	44.9	546	2 AAS04910	AAS04910 mycobacte

ALIGNMENTS

RESULT 1

ID	PUT2_HUMAN	STANDARD;	PRT;	563 AA.
AC	P30038; Q16882;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondrial precursor			
DE	(EC 1.5.1.12) (P5C dehydrogenase)			
GN	Name=ALDH4A1; Synonyms=P5CDH, ALDH4;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Kidney, and Retina;			
RX	MEDLINE=96199247; PubMed=8621661;			
RA	Hu C.-A., Lin W.-W., Valle D.;			
RT	"Cloning, characterization, and expression of cDNAs encoding human			
RL	delta 1-pyrroline-5-carboxylate dehydrogenase.";			
RN	J. Biol. Chem. 271:9795-9800(1996).			
RP	[2]			
RC	TISSUE=Liver;			
RX	MEDLINE=93009642; PubMed=1395511;			
RA	Hempel J., Eckey R., Romovacek H., Agarwal D.P.,			
RA	Goele H.W.;			
RT	"Human liver glutamic gamma-semialdehyde dehydrogenase: structural			
RT	relationship to the yeast enzyme.";			
RL	Comp. Biochem. Physiol. 102B:791-793(1992).			
RP	[3]			
RC	TISSUE=Liver;			
RX	MEDLINE=93162045; PubMed=1286669;			
RA	Hochstrasser D.F., Fruiger S., Paquet N., Bairoch A., Ravier F.,			
RA	Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,			
RA	Appel R.D., Hughes G.J.;			
RT	"Human liver protein map: a reference database established by			
RT	microsequencing and gel comparison.";			
RL	Electrophoresis 13:992-1001(1992).			
RN	[4]			
RP	VARIANT HP11 LEU-352, AND VARIANT LEU-16.			
RX	MEDLINE=93367029; PubMed=9700195;			
RA	Geraghty M.T., Vaughn D., Nicholson A.J., Lin W.-W.,			
RA	Jimenez-Sanchez G., Obie C., Flynn M.P., Valle D., Hu C.-A.;			
RT	"Mutations in the Delta1-pyrroline 5-carboxylate dehydrogenase gene			
RT	cause type II hyperprolinemia.";			
RL	Hum. Mol. Genet. 7:1411-1415(1998).			
CC	-!- FUNCTION: Irreversible conversion of delta-1-pyrroline-5-			
CC	carboxylate (P5C), derived either from proline or ornithine, to			
CC	glutamate. This is a necessary step in the pathway interconnecting			
CC	the urea and tricarboxylic acid cycles. The preferred substrate is			
CC	glutamic gamma-semialdehyde, other substrates include succinic,			

32	67	42.9	543	2	Q82JN1	Q82jnl streptomyc
33	67	42.9	546	2	Q8CJRI	Q8cjrli streptomyc
34	66	42.3	546	1	PUT2_AGABI	P78568 agaricus bi
35	61	39.1	581	2	Q95YV8	Q95tyv8 drosophila
36	61	39.1	608	2	Q8IN17	Q8ini7 drosophila
37	61	39.1	609	2	Q8IN18	Q8ini8 drosophila
38	61	39.1	614	2	Q8TQ08	Q8tqc8 drosophila
39	61	39.1	685	2	Q8LLB4	Q8lllb4 hordeum vul
40	61	39.1	685	2	Q8LLB6	Q8lllb6 hordeum vul
41	61	39.1	965	2	Q6HNO7	Q6hng7 bacillus th
42	61	39.1	965	2	Q73DU5	Q73du5 bacillus ce
43	61	39.1	965	2	Q81Y57	Q81ys7 bacillus an
44	61	39.1	965	2	AAS39550	Aas39550 bacillus
45	61	39.1	965	2	AAT29652	Aat29652 bacillus

CC Glutamic and adipic semialdehydes.
 CC -!- CATALYTIC ACTIVITY: L-pyrroline-5-carboxylate + NAD(+) + H(2)O =
 CC L-glutamate + NADH.
 CC -!- PATHWAY: Conversion from proline to glutamate; second step.
 CC -!- SUBUNIT: Homodimer.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -!- TISSUE SPECIFICITY: Highest expression is found in liver followed
 CC by skeletal muscle, kidney, heart, brain, placenta, lung and
 CC pancreas.
 CC -!- DISEASE: Defects in ALDH4A1 are the cause of hyperprolinemia type
 CC II (HPH) [MIM:239510]. HPH is characterized by the accumulation
 CC of delta-1-pyrroline-5-carboxylate (P5C) and proline. The disorder
 CC may be causally related to neurologic manifestations, including
 CC seizures and mental retardation.
 CC -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U24267; AAC50501.1; -.
 CC EMBL: U24266; AAC50500.1; -.
 CC SWISS-2DPAGE: P30038; HUMAN.
 CC OGP: P30038; -.
 CC Gene: HGNC:406; ALDH4A1.
 CC Reactome: P30038; -.
 CC MIN: 60811; -.
 CC MIN: 239510; -.
 CC DR GO: GO:0005759; C:mitochondrial matrix; TAS.
 CC DR GO: GO:0003842; F:1-pyrroline-5-carboxylate dehydrogenase act. . ; TAS.
 CC DR GO: GO:0004029; F:aldehyde dehydrogenase (NAD) activity; TAS.
 CC DR GO: GO:0005489; F:electron transporter activity; TAS.
 CC DR GO: GO:0006562; P:proline catabolism; TAS.
 CC DR InterPro: IPR002086; Aldehyde dehydr.
 CC DR Pfam: PF00171; Aldedh; 1.
 CC DR TIGRFAMs: TIGR01804; BADH; 1.
 CC DR PROSITE: PS00070; ALDEHYDE DEHYDR_CYS; 1.
 CC DR PROSITE: PS00687; ALDEHYDE DEHYDR_GLU; 1.
 CC KW Direct protein sequencing; Disease mutation; Mitochondrion; NAD;
 CC Oxidoreductase; Polymorphism; Proline metabolism; Transit peptide.
 CC FT TRANSIT 1 24 Mitochondrion.
 CC FT CHAIN 25 563 Delta-1-pyrroline-5-carboxylate
 CC dehydrogenase.
 CC FT NP_BIND 296 301 NAD (ADP part) (By similarity).
 CC FT ACT_SITE 314 314 By similarity.
 CC FT ACT_SITE 348 348 By similarity.
 CC FT VARIANT 16 16 P -> L (in allele ALDH4A1*4).
 CC FT VARIANT 352 352 S -> L (in HPH; allele ALDH4A1*3).
 CC FT CONFLICT 189 189 P -> L (in Ref. 2).
 CC FT CONFLICT 271 271 D -> E (in Ref. 2).
 CC SQ SEQUENCE 563 AA; 61751 MW; 8B864771B7DB5FF8 CRC64;
 CC
 CC Query Match 90.4%; Score 141; DB 1; Length 563;
 CC Best Local Similarity 93.3%; Pred. No. 9.4e-11;
 CC Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 GAVFSQDKVVQVQATKVLNRNADNFYINDR 31
 CC Db |||||
 CC 480 GAVFSQDKVVQVQATKVLNRNADNFYINDR 509
 CC
 CC RESULT 2
 CC Q81238 PRELIMINARY; PRT; 563 AA.
 CC AC Q81238;
 CC
 CC Query Match 90.4%; Score 141; DB 1; Length 563;
 CC Best Local Similarity 93.3%; Pred. No. 9.4e-11;
 CC Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 GAVFSQDKVVQVQATKVLNRNADNFYINDR 31
 CC Db |||||
 CC 480 GAVFSQDKVVQVQATKVLNRNADNFYINDR 509
 CC
 CC RESULT 3
 CC Q961F0 PRELIMINARY; PRT; 563 AA.
 CC AC Q961F0;
 CC DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 CC DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 CC DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 CC DE Aldehyde dehydrogenase 4A1.
 CC GN Name=ALDH4A1;
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE Aldehyde dehydrogenase 4A1.
 GN Name=ALDH4A1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Lung;
 RX MEDLINE=23388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC023600; AAH23600.1; -.
 DR GO: GO:0005759; C:mitochondrial matrix; IEA.
 DR GO: GO:0003842; F:1-pyrroline-5-carboxylate dehydrogenase act. . ; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR GO: GO:0006561; P:proline biosynthesis; IEA.
 DR InterPro: IPR002086; Aldehyde dehydr.
 DR Pfam: PF00171; Aldedh; 1.
 DR TIGRFAMs: TIGR01236; Dipy+5carboxyl; 1.
 DR PROSITE: PS00070; ALDEHYDE DEHYDR_CYS; UNKNOWN 1.
 DR PROSITE: PS00687; ALDEHYDE DEHYDR_GLU; UNKNOWN 1.
 SQ SEQUENCE 563 AA; 61719 MW; 4D964771B7DB5FFD CRC64;
 CC
 CC Query Match 90.4%; Score 141; DB 2; Length 563;
 CC Best Local Similarity 93.3%; Pred. No. 9.4e-11;
 CC Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 GAVFSQDKVVQVQATKVLNRNADNFYINDR 31
 CC Db |||||
 CC 480 GAVFSQDKVVQVQATKVLNRNADNFYINDR 509
 CC
 CC RESULT 3
 CC Q961F0 PRELIMINARY; PRT; 563 AA.
 CC AC Q961F0;
 CC DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 CC DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 CC DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 CC DE Aldehyde dehydrogenase 4A1.
 CC GN Name=ALDH4A1;
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC007581; AA07581.1; -;
 DR GO: GO:0005759; C:mitochondrial matrix; IEA.
 DR GO: GO:0003842; F:l-pyruvate-5-carboxylate dehydrogenase act. .; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR GO: GO:0006561; P:proline biosynthesis; IEA.
 DR InterPro: IPR002086; Aldehyde dehydr.
 DR Pfam: PF001171; Aldehyd; 1.
 DR TIGRFAMs: TIGR01236; Dipyr5carboxyl; 1.
 DR PROSITE: PS00687; ALDEHYDE DEHYDR. GLU; UNKNOWN 1.
 DR Query Match 90.4%; Score 141; DB 2; Length 563;
 Best Local Similarity 93.3%; Pred. No. 9.4e-11;
 Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 GAVFSQDKVQVQATKVLNRNADNFYINDR 31
 DB 480 GAVFSQDKVQVQATKVLNRNAGNFYINDK 509
 [1]
 RESULT 4
 Q7TP27 PRELIMINARY; PRT; 1465 AA.
 AC Q7TP27
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Bal-651.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Xu C.S., Li W.Q., Li Y.C., Wang L., Wang S.F., Han H.P., Wang G.P.,
 Chai L.Q., Yuan J.Y., Yang K.J., Yan H.M., Chang C.F., Zhao L.F.,
 Ma H., Shi J.B., Rahman S., Wang Q.N., Zhang J.B.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AV325227; AAP92628.1; -;
 DR GO: GO:0005759; C:mitochondrial matrix; IEA.

DR GO: GO:0003842; F:l-pyruvate-5-carboxylate dehydrogenase act. .; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0008152; P:metabolism; IEA.
 DR GO: GO:0006561; P:proline biosynthesis; IEA.
 DR InterPro: IPR002086; Aldehyde dehydr.
 DR InterPro: IPR005931; Dipyr5carboxyl.
 DR Pfam: PF001171; Aldehyd; 1.
 DR Pfam: PF01094; ANF receptor; 1.
 DR TIGRFAMs: TIGR01236; Dipyr5carboxyl; 1.
 DR PROSITE: PS00687; ALDEHYDE DEHYDR. GLU; UNKNOWN 1.
 DR Query Match 79.5%; Score 124; DB 2; Length 1465;
 Best Local Similarity 76.7%; Pred. No. 5.7e-08;
 Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 QY 2 GAVFSQDKVQVQATKVLNRNADNFYINDR 31
 DB 836 GAVFSQDKVQVQATKVLNRNAGNFYINDK 865
 [1]
 RESULT 5
 Q8RON1 PRELIMINARY; PRT; 381 AA.
 AC Q8RON1
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Aldh4a1 protein (Fragment).
 GN Name=A30035F14rik; Synonyms=Aldh4a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC026589; AAH26589.1; -;
 DR HSPF; P05091; 1002.
 DR MGD; MGI:2443883; A930035F14rik.
 DR GO: GO:0005739; C:mitochondrion; IEA.
 DR InterPro: IPR002086; Aldehyde dehydr.
 DR InterPro: IPR005931; Dipyr5carboxyl.
 DR Pfam: PF001171; Aldehyd; 1.
 DR TIGRFAMs: TIGR01236; Dipyr5carboxyl; 1.

DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; UNKNOWN 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN 1.
 FT NON TER 1
 SQ SEQUENCE 381 AA; 41755 MW; AC120FC168A5AC52 CRC64;

Query Match 78.8%; Score 123; DB 2; Length 381;
 Best Local Similarity 76.7%; Pred. No. 1.9e-08;
 Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAVFSQDKVVQVQATKVLRAADNFYINDR 31

Db 298 GAVFAQDKAIVQVQATRLRLNAAGNFYINDK 327

RESULT 6

ID Q8R1S2 PRELIMINARY; PRT; 549 AA.
 AC Q8R1S2;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Aldehyde protein (Fragment).
 GN Name=A930035F14Rik; Synonyms=Aldh4al;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Liver;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marisina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.,
 RA Krywinski M.I., Skalski U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Liver;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC024133; AAH24133.1; -;
 DR MGI; MGI:2443883; A930035F14Rik.
 DR GO; GO:0005739; C:mitochondrion; IDA.
 DR InterPro; IPSC02086; Aldehyde dehydr.
 DR InterPro; IPSC05931; Dipy5catbox1.
 DR Pfam; PF00171; Aldedh; 1.
 DR TIGRFAMs; TIGR01236; Dipy5catbox1; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; UNKNOWN 1.
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN 1.
 FT NON TER 1
 SQ SEQUENCE 549 AA; 60345 MW; B93CB793655D0A19 CRC64;

Query Match 78.8%; Score 123; DB 2; Length 549;
 Best Local Similarity 76.7%; Pred. No. 2.8e-08;
 Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAVFSQDKVVQVQATKVLRAADNFYINDR 31

Db 466 GAVFAQDKAIVQVQATRLRLNAAGNFYINDK 495

RESULT 7

ID Q8BXN3 PRELIMINARY; PRT; 562 AA.
 AC Q8BXN3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus adult retina cDNA, RIKEN full-length enriched library,
 DE clone:A930035F14 product:ALDEHYDE DEHYDROGENASE 4 FAMILY, MEMBER A1
 DE homolog.
 GN Name=A930035F14Rik;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RM Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RM Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RM Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kasaiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Retina;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Horii F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,

```
RA Nishio K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitho H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki Y., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK044712; EAC32045.1; -
DR MGD; MGI:2443883; A930035F14Rik.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR Pfam; PF00171; Aldedh; 1.
DR TIGRFAMs; TIGR01236; Dipy5carboxyl; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; UNKNOWN1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; UNKNOWN1.
SQ SEQUENCE 562 AA; 61752 MW; 0AFBF3FA72D5D1367 CRC64;

Query Match 78.8%; Score 123; DB 2; Length 562;
Best Local Similarity 76.7%; Pred. No. 2.9e-08;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAVFSQDKVQVQATKVLNRNADNFYINDR 31
DB 479 GAVFAQDKAIVQVATRLNRNAGNFYINDK 508

RESULT 8
Q8CHTO PRELIMINARY; PRT; 562 AA.
AC Q8CHTO
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Aldehyde dehydrogenase 4 family, member A1.
GN Name=Aldhd4l;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., McEwan P.J., McKernan K.J., Abramson R.D., Mullaly S.J.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Kidney;
RX Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
RA
```

```
DR EMBL; BC039281; AAH39281.2; -
DR EMBL; BC056226; AAH56226.1; -
DR MGD; MGI:2443883; A930035F14Rik.
DR GO; GO:0005739; C:mitochondrion; IDA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR005931; Dipy5carboxyl.
DR Pfam; PF00171; Aldedh; 1.
DR TIGRFAMs; TIGR01236; Dipy5carboxyl; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR_CYS; UNKNOWN1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR_GLU; UNKNOWN1.
SQ SEQUENCE 562 AA; 61810 MW; 0AFBF3FAA0C9C367 CRC64;

Query Match 78.8%; Score 123; DB 2; Length 562;
Best Local Similarity 76.7%; Pred. No. 2.9e-08;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAVFSQDKVQVQATKVLNRNADNFYINDR 31
DB 479 GAVFAQDKAIVQVATRLNRNAGNFYINDK 508

RESULT 9
Q7SY23 PRELIMINARY; PRT; 556 AA.
AC Q7SY23
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Similar to aldehyde dehydrogenase 4 family, member A1.
GN Name=zgc:63592;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., McEwan P.J., McKernan K.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RX Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC055155; AAH55155.1; -
DR GO; GO:0005739; C:mitochondrial matrix; IEA.
DR GO; GO:0003842; F:1-pyruvate-5-carboxylate dehydrogenase act. .; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0006561; P:proline biosynthesis; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR005931; Dipy5carboxyl.
DR
```

```

DR Pfam; PF00171; Aldehyd; 1.
DR TIGRFAMs; TIGR01236; Dipyrfscarboxyl; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; UNKNOWN 1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR GLU; UNKNOWN 1.
DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; UNKNOWN 1.
SQ SEQUENCE 556 AA; 61587 MW; 27973E9787CEA358 CRC64;

Query Match 71.2%; Score 111; DB 2; Length 556;
Best Local Similarity 65.7%; Pred. No. 1.3e-06;
Matches 20; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 GAVFSQDKVQVEATKVLNRNAADNFYINDR 31
Db |||||:|||||:|||||:|||||:|||||:
473 GAIFPQDKSVIEEAGKALRNAGNYINDK 502

RESULT 10
PUT2 SCHPO STANDARD; PRT; 548 AA.
AC 074766; P78880;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Probable delta-1-pyrroline-5-carboxylate dehydrogenase (EC 1.5.1.12)
DE (P5C dehydrogenase).
GN ORFNames=SPBC24C6.04;
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Ogell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McBride W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrall B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 189-548 FROM N.A.
RC STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomyces pombe
cDNA";
RL DNA Res. 4:363-369(1997).
CC -1- CATALYTIC ACTIVITY: 1-pyrroline-5-carboxylate + NAD(+) + H(2)O =
CC L-glutamate + NADH.
CC -1- PATHWAY: Conversion from proline to glutamate; second step.
CC -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; AL031786; CAA21148.1; -
DR EMBL; D89230; BAA13891.1; -
DR PIR; T39968; T39968.
DR GeneDB.SPombe; SPBC24C6.04; -
DR InterPro; IP002086; Aldehyde dehydr.
DR InterPro; IPR005931; Dipyrfscarboxyl.
DR Pfam; PF00171; Aldehyd; 1.
DR TIGRFAMs; TIGR01804; BADH; 1.
DR TIGRFAMs; TIGR01236; Dipyrfscarboxyl; 1.
DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; 1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR GLU; 1.
KW NAD, Oxidoreductase; Proline metabolism.
FT ACT_SITE 298 298 By similarity.
FT ACT_SITE 332 332 By similarity.
FT CONFLICT 210 210 N -> K (in Ref. 2).
FT CONFLICT 264 264 L -> F (in Ref. 2).
FT CONFLICT 273 274 FR -> IC (in Ref. 2).
FT CONFLICT 294 294 K -> N (in Ref. 2).
FT CONFLICT 502 502 E -> G (in Ref. 2).
FT CONFLICT 516 516 K -> E (in Ref. 2).
FT CONFLICT 526 526 F -> Y (in Ref. 2).
FT CONFLICT 531 531 S -> Y (in Ref. 2).
FT CONFLICT 536 536 F -> L (in Ref. 2).
SQ SEQUENCE 548 AA; 60218 MW; E14D37E32A0B4B6E CRC64;

Query Match 61.5%; Score 96; DB 1; Length 548;
Best Local Similarity 60.0%; Pred. No. 0.00015;
Matches 18; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 GAVFSQDKVQVEATKVLNRNAADNFYINDR 31
Db |||||:|||||:|||||:|||||:|||||:
465 GSIFPAQDRVVRVKLTDLNRNAAGNFYINDK 494

RESULT 11
Q9YDW2 PRELIMINARY; PRT; 553 AA.
AC Q9YDW2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 553aa long hypothetical delta-1-pyrroline-5-carboxylate
DE dehydrogenase.
GN OrderedLocustNames=APE0807;
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310139; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79785.1; -
DR PIR; A72673; A72673.
DR GO; GO:0005759; C:mitochondrial matrix; IEA.

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Delta-1-pyrroline-5-carboxylate dehydrogenase.
 GN ORFNamesChr3.0160;
 OS Leishmania major.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Friedlin;
 RX MEDLINE=22738071; PubMed=12853638;
 RA Worthey E.A., Martinez-Caivillo S., Schnauffer A., Aggarwal G.,
 RA Cawthra J., Fazelinia G., Fong C., Fu G., Hassebrock M., Hixson G.,
 RA Ivers A.C., Kiser P., Marsolini F., Rickell E., Salavati R., Sisk E.,
 RA Sunkin S.M., Stuart K.D., Myler P.J.;
 RT "Leishmania major chromosome 3 contains two long convergent
 RT polycistronic gene clusters separated by a tRNA gene.";
 RL Nucleic Acids Res. 31:4201-4210(2003).
 DR EMBL; ACL25735; AAME8988.1; ..
 DR GO; GO:0005759; C:mitochondrial matrix; IEA.
 DR GO; GO:0003842; F:1-pyrroline-5-carboxylate dehydrogenase act. .; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR GO; GO:0006561; P:proline biosynthesis; IEA.
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR InterPro; IPR005931; Dipyrscarboxyl.
 DR Pfam; PF00171; Aldehyd; 1.
 DR TIGRfam; TIGR01236; Dipyrscarboxyl; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; UNKNOWN 1.
 DR PROSITE; PS00687; ALDEHYDE DEHYDR GLU; UNKNOWN 1.
 SQ SEQUENCE 560 AA; 61899 MW; 30C9A49DEE598535 CRC64;

Query Match 56.4%; Score 88; DB 2; Length 560;

Best Local Similarity 53.3%; Pred. No. 0.002;

Matches 16; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 2 GAVFSQDKVVQVQATKVLRNAADNFYINDR 31

DB 471 GAVFSREARIRADKYLRYAGNYVNDK 500

RESULT 15

ID Q75E29 PRELIMINARY; PRT; 573 AA.
 AC Q75E29;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE AAL071Cp.
 GN Name=AAL071C;
 OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
 OX NCBI_TaxID=33169;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 10895;
 RX PubMed=15001715;
 RA Dietrich F.S., Voegeli S., Brachat S., Lerch A., Gates K., Steiner S.,
 RA Mohr C., Pohlmann R., Luedi P., Choi S., Wing R.A., Flavler A.,
 RA Gaffney T.D., Philippsen P.;
 RT "The Ashbya gossypii genome as a tool for mapping the ancient
 RT Saccharomyces cerevisiae genome.";
 RL Science 304:304-307(2004).
 DR EMBL; AE016880; AAS50295.1; ..
 DR AGD; AAL071C; ..
 DR InterPro; IPR002086; Aldehyde dehydr.
 DR InterPro; IPR005931; Dipyrscarboxyl.
 DR Pfam; PF00171; Aldehyd; 1.
 DR TIGRfam; TIGR01236; Dipyrscarboxyl; 1.
 DR PROSITE; PS00070; ALDEHYDE DEHYDR CYS; UNKNOWN 1.
 DR PROSITE; PS00687; ALDEHYDE DEHYDR GLU; UNKNOWN 1.
 SQ SEQUENCE 573 AA; 62967 MW; 300B396AB343C0C2 CRC64;

Query Match 55.8%; Score 87; DB 2; Length 573;
 Best Local Similarity 53.3%; Pred. No. 0.0028;

Matches 16; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

OY 2 GAVFSQDKVVQVQATKVLRNAADNFYINDR 31

DB 489 GAVFARDRDAIAAADRLKLYAAGNFYINDK 518

Search completed: November 10, 2004, 12:27:08

JOB time : 36.9182 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 9.8287 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-14

Perfect score: 171

Sequence: 1 TGTGAPRFKEVQELNSALHQSLLDIYRTLHP 33

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/protdata/1/iaa/5A.COMB.pep.*
- 2: /cgn2_6/protdata/1/iaa/5B.COMB.pep.*
- 3: /cgn2_6/protdata/1/iaa/6A.COMB.pep.*
- 4: /cgn2_6/protdata/1/iaa/6B.COMB.pep.*
- 5: /cgn2_6/protdata/1/iaa/PGTUS.COMB.pep.*
- 6: /cgn2_6/protdata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	34.5	511	4	US-09-252-991A-26528
2	54.5	31.9	294	2	US-08-923-856-3
3	54.5	31.9	294	3	US-09-216-294-3
4	53	31.0	815	4	US-09-538-092-1196
5	51	29.8	224	4	US-09-248-796A-20686
6	51	29.8	312	4	US-09-710-279-1210
7	51	29.8	323	3	US-09-134-001C-3789
8	51	29.8	560	3	US-09-134-001C-4343
9	51	29.8	584	4	US-09-710-279-2832
10	50.5	29.5	462	4	US-09-520-822A-4
11	50	29.2	219	4	US-09-270-767-34504
12	50	29.2	219	4	US-09-270-767-49721
13	49	28.7	241	4	US-09-710-279-856
14	49	28.7	249	3	US-09-134-001C-3351
15	49	28.7	260	4	US-09-540-236-2594
16	49	28.7	346	4	US-09-187-906-15
17	49	28.7	363	4	US-09-543-681A-5494
18	49	28.7	397	3	US-09-220-528-64
19	49	28.7	397	4	US-09-187-906-17
20	49	28.7	538	4	US-09-252-991A-23068
21	49	28.7	674	1	US-08-803-973-7
22	49	28.7	674	1	US-08-803-972-7
23	49	28.7	707	1	US-08-803-973-12
24	49	28.7	707	1	US-08-803-972-12
25	49	28.7	1456	1	US-08-803-972-2
26	49	28.7	1456	1	US-08-803-972-2
27	48.5	28.4	271	3	US-09-318-754A-2

ALIGNMENTS

RESULT 1

US-09-252-991A-26528
; Sequence 26528, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 26528

; LENGTH: 511

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26528

Query Match 34.5%; Score 59; DB 4; Length 511;

Best Local Similarity 48.0%; Pred. No. 2;

Matches 12; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Query 1 TGTGAPRFKEVQELNSALHQSLLDI 25

Db 241 TVTGVPTGLKDLDELSSGLQKSDLI 265

RESULT 2

US-08-923-856-3

; Sequence 3, Application US/08923856

; Patent No. 5928894

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; APPLICANT: Tang, Tom

; APPLICANT: Corley, Neil C.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: HUMAN ACTIVA-ORF4-LIKE PROTEIN

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

Sequence 4, Appli
Sequence 1, Appli
Sequence 45, Appli
Sequence 13, Appli
Sequence 5, Appli
Sequence 1, Appli
Sequence 31538, A
Sequence 25971, A
Sequence 4822, Ap
Sequence 6333, Ap
Sequence 10798, A
Sequence 6448, Ap
Sequence 6296, Ap
Sequence 526, App
Sequence 353, App
Sequence 2380, Ap
Sequence 7771, Ap
Sequence 68, Appli

28 48.5 28.4 271 3 US-09-318-733A-4
29 48 28.1 46 3 US-09-124-671-1
30 48 28.1 64 3 US-09-091-814-45
31 48 28.1 115 3 US-09-124-671-13
32 48 28.1 248 2 US-08-805-965-5
33 48 28.1 262 2 US-08-805-965-1
34 48 28.1 279 4 US-09-252-991A-31538
35 48 28.1 302 4 US-09-252-991A-25971
36 48 28.1 428 4 US-09-583-110-4822
37 47.5 27.8 257 4 US-09-134-000C-6333
38 47 27.5 114 4 US-09-489-039A-10798
39 47 27.5 118 4 US-09-513-999C-6448
40 47 27.5 140 4 US-09-621-976-6296
41 47 27.5 174 4 US-09-149-476-526
42 47 27.5 234 4 US-09-149-476-353
43 47 27.5 281 4 US-09-540-236-2380
44 47 27.5 296 4 US-09-543-681A-7771
45 47 27.5 426 4 US-09-602-787A-68

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 46816
;
US-09-216-294-3

Query Match          31.9%; Score 54.5; DB 3; Length 294;
Best Local Similarity 52.2%; Pred. No. 4.9;
Matches 12; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY      4 GAPRIKEVQELNSALHQSOLDID 26
       |||:||||:|||||||
Db      40 GAPK-AKETLGLASHTGDDLE 61

RESULT 4
US-09-538-092-1196
; Sequence 1196, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSegFormatter Version 0.9

```

```

; SEQUENCE INFORMATION
; SEQ ID NO 1196
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P52756
; US-09-538-092-1196

Query Match      31.0%; Score 53; DB 4; Length 815;
Best Local Similarity 35.1%; Pred. No. 27;
Matches 13; Conservative       7; Mismatches 7; Indels 10; Gaps 1;

QY    7 RRIKEVQ-----ELNSALMQSDLDIYRTHLP 33
      |||:| | :|||:| |:|:|:|
DB    263 RLIKDKQTQQNRGFAFVQLSSAMDASQLLQLSLHP 299

RESULT 5
US-09-248-796A-20686
; Sequence 20686, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20686
```



```

; LENGTH: 224
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20686

Query Match          29.8%; Score 51; DB 4; Length 224;
Best Local Similarity 55.6%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY      16 NSALHOSDLIDIYRTLHP 33
      ||| :||| :|||
Db      41 NSHLTHADLIKFYQTCHP 58

RESULT 6
US-09-710-279-1210
; Sequence 1210, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMWERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1210
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-1210

Query Match          29.8%; Score 51; DB 4; Length 312;
Best Local Similarity 47.8%; Pred. No. 17;
Matches 11; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY      11 EVQELNSALHQSDLDIYRTLHP 33
      : : : : :||| :||| :|||
Db      4 KMMEIMEALEQSELIITRRHLRP 26

RESULT 7
US-09-134-001C-3789
; Sequence 3789, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3789
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3789

Query Match          29.8%; Score 51; DB 3; Length 323;
Best Local Similarity 47.8%; Pred. No. 18;
Matches 11; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY      11 EVQELNSALHQSDLDIYRTLHP 33

```

; APPLICANT: Chen, Xiaojiang
; TITLE OF INVENTION: Vaccine compositions and methods for human papilloma
; FILE REFERENCE: Harvard/Harrison 12687/1120
; CURRENT APPLICATION NUMBER: US/09/520,822A
; CURRENT FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/125208
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/148544
; PRIOR FILING DATE: 1999-08-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Human papillomavirus type 18
US-09-520-822A-4

Query Match 29.5%; Score 50.5; DB 4; Length 462;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 12; Conservative 4; Mismatches 7; Indels 1; Gaps 1;
QY 5 APRKEVQELNSALHQSLLDIY 28
DB 329 SPEYI-ELQPLVSATEDNDUFDIY 351

RESULT 11
US-09-270-767-34504
; Sequence 34504, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 34504
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-34504

Query Match 29.2%; Score 50; DB 4; Length 219;
Best Local Similarity 33.3%; Pred. No. 16;
Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;
QY 1 TGTGAPRIKEVQELNSALHQSLL 24
DB 110 TGXGSPRFVEKXSASESVTYADI 133

RESULT 12
US-09-270-767-49721
; Sequence 49721, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49721
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:

; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-49721

Query Match 29.2%; Score 50; DB 4; Length 219;
Best Local Similarity 33.3%; Pred. No. 16;
Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;
QY 1 TGTGAPRIKEVQELNSALHQSLL 24
DB 110 TGXGSPRFVEKXSASESVTYADI 133

RESULT 13
US-09-710-279-856
; Sequence 856, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 856
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-856

Query Match 28.7%; Score 49; DB 4; Length 241;
Best Local Similarity 40.0%; Pred. No. 25;
Matches 10; Conservative 5; Mismatches 8; Indels 2; Gaps 1;
QY 11 EVQELNSAL--HQSDLDIYRTLHP 33
DB 86 KLNDLNDRLTWHEHMDLKDVTQTPQ 110

RESULT 14
US-09-134-001C-3351
; Sequence 3351, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3351
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3351

Query Match 28.7%; Score 49; DB 3; Length 249;
Best Local Similarity 40.0%; Pred. No. 26;
Matches 10; Conservative 5; Mismatches 8; Indels 2; Gaps 1;
QY 11 EVQELNSAL--HQSDLDIYRTLHP 33
DB 94 KLNDLNDRLTWHEHMDLKDVTQTPQ 118

Search completed: November 10, 2004, 12:32:27
Job time : 9.8287 secs

This Page Blank (uspto)

; APPLICANT: Sun, Yongming
 ; APPLICANT: Recipon, Hervé
 ; APPLICANT: Chen, Sei-Yu
 ; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
 ; FILE REFERENCE: DEX-0265
 ; CURRENT APPLICATION NUMBER: US/10/012,600B
 ; CURRENT FILING DATE: 2001-11-06
 ; PRIOR APPLICATION NUMBER: US 60/246,109
 ; PRIOR FILING DATE: 2000-11-06
 ; NUMBER OF SEQ ID NOS: 240
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 221
 ; LENGTH: 363
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-012-600B-221

Query Match 73.4%; Score 125.5; DB 14; Length 363;
 Best Local Similarity 42.6%; Pred. No. 7.1e-10;
 Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;
 QY 3 TGAPRFYKEV-----QELNSALHQS DLI 25
 DB 118 TGAPRFYKEVLDLQRLNSHTIIVGDFNTLLSTLDRSMQKVNKDTQELNSALHQS DLI 177
 QY 26 DIYRTLHP 33
 DB 178 DIYRTLHP 185

RESULT 3
 US-10-082-828A-257
 ; Sequence 257, Application US/10082828A
 ; Publication No. US20030175715A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sun, Yongming
 ; APPLICANT: Recipon, Hervé
 ; APPLICANT: Salceda, Susana
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Turner, Leah
 ; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
 ; FILE REFERENCE: DEX-0247
 ; CURRENT APPLICATION NUMBER: US/10/082,828A
 ; CURRENT FILING DATE: 2002-07-09
 ; PRIOR APPLICATION NUMBER: US 60/243,805
 ; PRIOR FILING DATE: 2000-10-27
 ; NUMBER OF SEQ ID NOS: 266
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 257
 ; LENGTH: 1031
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-082-828A-257

Query Match 73.4%; Score 125.5; DB 14; Length 1031;
 Best Local Similarity 42.6%; Pred. No. 2.5e-09;
 Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;
 QY 3 TGAPRFYKEV-----QELNSALHQS DLI 25
 DB 21 TGAPRFYKEVLDLQRLDSHTLMDFTPLSTLDRSTRQKVNKDTQELNSALHQS DLI 80
 QY 26 DIYRTLHP 33
 DB 81 DIYRTLHP 88

RESULT 4
 US-10-025-201-3
 ; Sequence 3, Application US/10025201
 ; Publication No. US20030003468A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Crow, Mary K.

; TITLE OF INVENTION: MARKERS FOR DISEASE SUSCEPTIBILITY AND TARGETS FOR THERAPY
 ; FILE REFERENCE: 5983/2H567
 ; CURRENT APPLICATION NUMBER: US/10/025,201
 ; CURRENT FILING DATE: 2001-12-19
 ; PRIOR APPLICATION NUMBER: 60/256,673
 ; PRIOR FILING DATE: 2000-12-19
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 1275
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: GenBank Accession No. US20030003468A1 U09116
 ; DATABASE ENTRY DATE: 1995-02-02
 ; RELEVANT RESIDUES: (1)..(1275)
 ; US-10-025-201-3

Query Match 73.4%; Score 125.5; DB 14; Length 1275;
 Best Local Similarity 42.6%; Pred. No. 3.2e-09;
 Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;
 QY 3 TGAPRFYKEV-----QELNSALHQS DLI 25
 DB 119 TGAPRFYKEVLDLQRLDSHTLMDFTPLSTLDRSTRQKVNKDTQELNSALHQS DLI 178
 QY 26 DIYRTLHP 33
 DB 179 DIYRTLHP 186

RESULT 5
 US-10-114-270-30
 ; Sequence 30, Application US/10114270
 ; Publication No. US20040030110A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Guo, Xiaojia
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: Miller, Charles E.
 ; APPLICANT: Malyankar, Uriel M.
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Liu, Ziaohong
 ; APPLICANT: Gusev, Vladimir Y.
 ; APPLICANT: Li, Li
 ; APPLICANT: Vernet, Corine
 ; APPLICANT: Zerhusen, Bryan D.
 ; APPLICANT: Gorman, Linda
 ; APPLICANT: Shenoy, Suresh G.
 ; APPLICANT: Pena, Carol E.A.
 ; APPLICANT: Smithson, Glennda
 ; APPLICANT: Burgess, Catherine E.
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Shimkets, Richard A.
 ; APPLICANT: Gangolli, Esha A.
 ; APPLICANT: Taupier Jr., Raymond J.
 ; APPLICANT: Casman, Stacie J.
 ; APPLICANT: Ji, Weizhen
 ; APPLICANT: Anderson, David W.
 ; APPLICANT: Liette, Mario W.
 ; APPLICANT: Rastelli, Luca
 ; APPLICANT: Edinger, Shlomit R.
 ; APPLICANT: Stone, David J.
 ; APPLICANT: MacDougall, John R.
 ; APPLICANT: Rothenberg, Mark E.
 ; TITLE OF INVENTION: No. US20040030110A1el Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-322C
 ; CURRENT APPLICATION NUMBER: US/10/114,270
 ; CURRENT FILING DATE: 2002-11-27
 ; PRIOR APPLICATION NUMBER: 60/281,086
 ; PRIOR FILING DATE: 2001-04-03
 ; PRIOR APPLICATION NUMBER: 60/281,136

```
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,863
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/281,906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282,020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/282,930
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,934
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283,512
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/283,710
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/284,234
; PRIOR FILING DATE: 2001-04-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 470
; SEQ ID NO 30
; LENGTH: 1272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-114-270-30

Query Match      68.1%; Score 116.5; DB 15; Length 1272;
Best Local Similarity 41.2%; Pred. No. 6.8e-08;
Matches 28; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

QY 3 TGAPRIKEV-----QELNSALHQSGLI 25
Db 115 TGAPTIKQVLDRLDLSHTLIMGDFNPLSTLDRSTRQKYNKDTQELNSALHQSGLI 174
QY 26 DIYRTLHP 33
Db 175 DIYRTLHP 182

RESULT 6
US-10-243-552-850
; Sequence 850, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yuning
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10/243,552
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
```

```
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 850
; LENGTH: 936
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-850

Query Match      67.8%; Score 116; DB 14; Length 936;
Best Local Similarity 87.5%; Pred. No. 5.6e-08;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 10 KEVQELNSALHQSGLIDIYRTLHP 33
Db 21 KDIQELNSALHQAQDLIDIYRTLHP 44

RESULT 7
US-10-415-615-3
; Sequence 3, Application US/10415615
; Publication No. US20040101943A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: YUE, Henry
; APPLICANT: LU, Yan
; APPLICANT: DING, Li
; APPLICANT: TANG, Y. Tom
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: LAL, Preeti G.
; TITLE OF INVENTION: NUCLEIC ACID MODIFICATION ENZYMES
; FILE REFERENCE: PI-0280 USN
; CURRENT APPLICATION NUMBER: US/10/415,615
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: PCT/US01/46301
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/245,458
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/255,107
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 1258887CD1
US-10-415-615-3

Query Match      66.4%; Score 113.5; DB 16; Length 1274;
Best Local Similarity 38.2%; Pred. No. 1.9e-07;
Matches 26; Conservative 3; Mismatches 2; Indels 37; Gaps 1;

QY 3 TGAPRIKEV-----QELNSALHQSGLI 25
Db 119 TGVPRFIKQVLDRLDLSHTLIMGDFNPLSTLDRSMROKFNKDIQKLSALHQAQDLI 178
QY 26 DIYRTLHP 33
Db 179 DIYRTLHP 186

RESULT 8
US-10-104-047-2288
; Sequence 2288, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
```



```

; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31728
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL117259.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 15
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 8.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
; OTHER INFORMATION: SWISSPROT HIT: P08547, EVALUE 2.00e-12
US-10-029-386-31728
```

```

Query Match      42.4%; Score 72.5; DB 14; Length 110;
Best Local Similarity 54.2%; Pred. NO. 0.012;
Matches 13; Conservative 8; Mismatches 2; Indels 1; Gaps 1;
```

```

Qy 10 KEVQELNSALHQSGLDIYRTLHP 33
Db 59 KDIIEKUNIQ-DLWDIYRTLHP 81
```

```

RESULT 12
US-09-864-408A-7726
; Sequence 7726, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7726
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (16)..(16)
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (22)..(22)
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (104)..(104)
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (111)..(111)
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid
US-09-864-408A-7726
```

```

Query Match      40.4%; Score 69; DB 11; Length 114;
Best Local Similarity 65.8%; Pred. NO. 0.041;
Matches 13; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
```

```

Qy 14 ELNSALHQSGLDIYRTLHP 33
Db 18 DLNGXTDQLDLDIYRTLHP 37
```

```

RESULT 13
US-10-108-260A-4488
; Sequence 4488, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4488
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4488
```

```

Query Match      39.2%; Score 67; DB 15; Length 130;
Best Local Similarity 68.4%; Pred. NO. 0.094;
Matches 13; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```

Qy 15 LNSALHQSGLDIYRTLHP 33
Db 3 LNYTLQMDLTDIYRTLHP 21
```

```

RESULT 14
US-10-000-256A-143
; Sequence 143, Application US/10000256A
; Publication No. US20030039983A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
; FILE REFERENCE: DEX-0259
; CURRENT APPLICATION NUMBER: US/10/000,256A
; CURRENT FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: 60/244,782
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 143
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-000-256A-143
```

```

Query Match      36.8%; Score 63; DB 14; Length 230;
Best Local Similarity 52.2%; Pred. NO. 0.73;
Matches 12; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
```

```

Qy 10 KEVQELNSALHQSGLDIYRTLH 32
Db 11 KKTLDLNLATLDQMDLTDIYRTVY 33
```

```

RESULT 15
US-09-864-408A-8046
; Sequence 8046, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1el Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
```

```

; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8046
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (11)-(11)
; OTHER INFORMATION: wherein Xaa may be any naturally occurring amino acid
US-09-864-408A-8046

```

```

Query Match      35.1%; Score 60; DB 11; Length 99;
Best Local Similarity 44.4%; Pred. No. 0.74;
Matches 12; Conservative 9; Mismatches 4; Indels 2; Gaps 1;

```

```

Qy      9 IKEV--CELSALHQSGLDIDYRTLHP 33
      ||:::||||:||||:||||:|
Db      70 IKIRTEELANTINQTYLISLYRILYP 96

```

Search completed: November 10, 2004, 16:35:58
Job time : 40.8287 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:36:51 ; Search time 6.6713 seconds
(without alignments)
475.942 Million cell updates/sec

Title: US-10-092-750-14

Perfect score: 171

Sequence: 1 TGTGAPRFKEVQELNSALHQSLDIDIVRTLHP 33

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125.5	73.4	1259	4 GNHUL1	retrovirus-related
2	125.5	73.4	1275	2 B28096	line-1 protein ORF
3	125.5	73.4	1275	2 I38588	reverse transcript
4	125.5	73.4	1275	2 S65824	reverse transcript
5	125.5	73.4	1280	2 B34087	hypothetical prote
6	78	45.6	1260	4 GNLR11	retrovirus-related
7	62	36.3	202	2 S21346	probable pol polyp
8	60.5	35.4	500	2 S16788	probable reverse t
9	57	33.3	1281	1 GNMSLL	retrovirus-related
10	55	32.2	407	2 F69805	hypothetical prote
11	54.5	31.9	234	2 S18542	hypothetical prote
12	54	31.6	501	2 T21973	hypothetical prote
13	53.5	31.3	178	2 S77950	probable histidine
14	52.5	30.7	448	2 E70798	hypothetical prote
15	52	30.4	433	2 S36565	L2 protein - human
16	51.5	30.1	518	2 D71288	probable glutamate
17	51	29.8	823	2 H72282	maltose ABC transp
18	51	29.8	944	2 T38130	probable helicase
19	50.5	29.5	292	2 F82385	hypothetical prote
20	50.5	29.5	462	1 P20W18	L2 protein - human
21	50.5	29.5	579	2 F85311	flagellar biosynth
22	50.5	29.5	579	2 F64747	membrane protein f
23	50.5	29.5	918	2 S23377	collagen alpha 2(V
24	50.5	29.5	1022	2 S04111	collagen alpha 2(V
25	50.5	29.2	111	2 A81387	hypothetical prote
26	50	29.2	710	2 T50469	hypothetical prote
27	50	29.2	710	2 A25657	hypothetical prote
28	50	29.2	710	2 T50475	hypothetical prote
29	50	29.2	720	2 T50475	hypothetical prote

30 49 28.7 397 2 JE0082
31 49 28.7 527 2 A83453
32 49 28.7 755 2 A44315
33 49 28.7 1456 1 KXWGPV
34 48.5 28.4 271 2 T47119
35 48.5 28.4 407 2 T48739
36 48 28.1 127 2 F89858
37 48 28.1 240 2 T09700
38 48 28.1 270 2 H83327
39 48 28.1 283 2 C72492
40 48 28.1 355 2 B81257
41 48 28.1 369 2 F81674
42 48 28.1 512 2 G86773
43 48 28.1 1551 2 T18941
44 48 28.1 1958 2 T39808
45 47.5 27.8 386 2 H84522

ALIGNMENTS

RESULT 1

GNHUL1

retrovirus-related reverse transcriptase pseudogene - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1988 #sequence_revision 04-Jan-1996 #text_change 09-Jul-2004

C:Accession: A25313

R:Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.

Nature 321, 625-628, 1986

A:Title: L1 family of repetitive DNA sequences in primates may be derived from a sequence

A:Reference number: A93381; MUID:86230917; PMID:2423883

A:Accession: A25313

A>Status: conceptual translation of pseudogene

A:Molecule type: DNA

A:Residues: 1-1259 <HAT>

A:Cross-references: UNIPROT:P08547

A>Note: this sequence was constructed from an alignment of published and unpublished seq

C:Keywords: reverse transcriptase; pseudogene

Query Match 73.4%; Score 125.5; DB 4; Length 1259;

Best Local Similarity 42.6%; Pred. No. 6.4e-09;

Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY

3 TGAPRFKEV-----QELNSALHQSLDI 25

|||||

118 TGAPRFKEVSDIQRDLSDTLDSTROKINKDQELNSALHQADLI 177

|||||

QY 26 DIYRTLHP 33

|||||

Db 178 DIYRTLHP 185

RESULT 2

B28096

line-1 protein ORF2 - human

C:Species: Homo sapiens (man)

C:Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 09-Jul-2004

C:Accession: B28096

R:Skowronski, J.; Fanning, T.G.; Singer, M.F.

Mol. Cell. Biol. 8, 1385-1397, 1988

A:Title: Unit-length line-1 transcripts in human teratocarcinoma cells.

A:Reference number: A28096; MUID:88245405; PMID:2454389

A:Accession: B28096

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1275 <SKO>

A:Cross-references: UNIPROT:Q15604; UNIPROT:Q9UN80; UNIPROT:Q12881; UNIPROT:O00363; UNIPR

PROT:Q9YSK0; UNIPROT:O00366; UNIPROT:Q8TE30; UNIPROT:O00375

C:Superfamily: pol polypotein

Query Match 73.4%; Score 125.5; DB 2; Length 1275;

Best Local Similarity 42.6%; Pred. No. 6.5e-09;

Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

```

Qy 3 TGAPRFKEV-----QELNSALHQSGLI 25
Db 119 TGAPRFKQVLSLQRLDLSHTLIMGFNTPLSTLRSTRQKYNKDTQELNSALHQSGLI 178

Qy 26 DIYRTLHP 33
Db 179 DIYRTLHP 186

RESULT 3
I38588
reverse transcriptase homolog - human retrotransposon L1
N:Alternate names: ORF2 protein
C:Species: Homo sapiens (man)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I38588
R:Holmes, S.E.; Dombroski, B.A.; Krebs, C.M.; Boehm, C.D.; Kazazian, H.H.
Nature Genet. 7, 143-148, 1994
A>Title: A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q
A:Reference number: I38587; MUID:95004577; PMID:7920631
A:Accession: I38588
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1275 <RES>
A:Cross-references: UNIPROT:Q12881; EMBL:M80340; NID:g339767; PIDN:AAA51622.1; PID:g3397
C:Superfamily: pol polyprotein

Query Match 73.4%; Score 125.5; DB 2; Length 1275;
Best Local Similarity 42.6%; Pred. No. 6.5e-09;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

Qy 3 TGAPRFKEV-----QELNSALHQSGLI 25
Db 119 TGAPRFKQVLSLQRLDLSHTLIMGFNTPLSTLRSTRQKYNKDTQELNSALHQSGLI 178

Qy 26 DIYRTLHP 33
Db 179 DIYRTLHP 186

RESULT 4
S65824
reverse transcriptase homolog - human transposon L1.1
C:Species: Homo sapiens (man)
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S65824
R:Dombroski, B.A.
submitted to the EMBL Data Library, January 1992
A:Description: Isolation of an active human transposable element.
A:Reference number: S65823
A:Accession: S65824
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1275 <DOM>
A:Cross-references: UNIPROT:Q15604; EMBL:M80340; NID:g339767; PIDN:AAA51622.1; PID:g3397
C:Superfamily: pol polyprotein

Query Match 73.4%; Score 125.5; DB 2; Length 1275;
Best Local Similarity 42.6%; Pred. No. 6.5e-09;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

Qy 3 TGAPRFKEV-----QELNSALHQSGLI 25
Db 119 TGAPRFKQVLSLQRLDLSHTLIMGFNTPLSTLRSTRQKYNKDTQELNSALHQSGLI 178

Qy 26 DIYRTLHP 33
Db 179 DIYRTLHP 186

RESULT 5
B34087

```

```

hypothetical protein (L1H 3' region) - human
C:Species: Homo sapiens (man)
C>Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 09-Jul-2004
C:Accession: B34087
R:Scott, A.F.; Schmeckpeper, B.J.; Abdelrazik, M.; Comey, C.T.; O'Hara, B.; Rossiter, J.I.
Genomics 1, 113-125, 1987
A>Title: Origin of the human L1 elements: proposed progenitor genes deduced from a conserved
A:Reference number: A34087; MUID:88085185; PMID:3692483
A:Accession: B34087
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1280 <SCO>
A:Cross-references: UNIPROT:Q9Y5K0
C:Superfamily: pol polyprotein

Query Match 73.4%; Score 125.5; DB 2; Length 1280;
Best Local Similarity 42.6%; Pred. No. 6.5e-09;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

Qy 3 TGAPRFKEV-----QELNSALHQSGLI 25
Db 124 TGAPRFKQVLSLQRLDLSHTLIMGFNTPLSTLRSTRQKYNKDTQELNSALHQSGLI 183

Qy 26 DIYRTLHP 33
Db 184 DIYRTLHP 191

RESULT 6
GNLRLL
retrovirus-related reverse transcriptase pseudogene - slow loris
C:Species: Nycticebus coucang (slow loris)
C>Date: 31-Mar-1988 #sequence_revision 04-Jan-1996 #text_change 09-Jul-2004
C:Accession: B25313
R:Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.
Nature 321, 625-628, 1986
A>Title: L1 family of repetitive DNA sequences in primates may be derived from a sequence
A:Reference number: A93381; MUID:86230917; PMID:2423883
A:Accession: B25313
A>Status: conceptual translation of pseudogene
A:Molecule type: DNA
A:Residues: 1-1260 <HAT>
A:Cross-references: UNIPROT:P08548
A>Note: this sequence was constructed from an alignment of six sequences, determined by t
C:Keywords: reverse transcriptase; pseudogene

Query Match 45.6%; Score 78; DB 4; Length 1260;
Best Local Similarity 58.3%; Pred. No. 0.024; Mismatches 7; Indels 0; Gaps 0;
Matches 14; Conservative 3;

Qy 10 KEVQELNSALHQSGLI DIYRTLHP 33
Db 162 KSILDLSNSTIQLHDLT DIYRTLHP 185

RESULT 7
S21346
probable pol polyprotein-related protein 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C:Accession: S21346
R:Schmittz, E.; Mohr, E.
submitted to the EMBL Data Library, June 1990
A:Reference number: S21345
A:Accession: S21346
A:Molecule type: DNA
A:Residues: 1-202 <SCH>
A:Cross-references: UNIPROT:Q63304; EMBL:X53581; NID:g56586; PIDN:CAA37645.1; PID:g56588
A:Experimental source: strain Wistar
C:Superfamily: pol polyprotein
C:Keywords: polypolyprotein

Query Match 36.3%; Score 62; DB 2; Length 202;

```

A;Accession: I49130
A;Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A;Residues: 1-85,'L','R','78'-358,'K','360-706','F','708-735','A','737-760','W','762-927','D','929-1281'
A;Cross-references: EMBI:UI15647; NID:g558906; PIDN:AAA67727.1; PID:g558908
R;Mottez, B.; Rogan, P.K.; Manuelidis, L.
Nucleic Acids Res 14, 3119-3136, 1986
A>Title: Conservation in the 5' region of the long interspersed mouse L1 repeat: implicat
A;Reference number: A23772; MUID:86178799; PMID:33008107
A;Accession: A23772
A:Molecule type: DNA
A;Residues: 'NNQSNHGHTNQKEDSHKNR',1-245,'K','247-423','SYTQQNKWTWKTN',439,'WTDTRYQS' <MO1
A;Cross-references: GB:X03725; NID:g52829; PIDN:CAA27363.1; PID:g1334115
C;Superfamily: pol polyprotein
C;Keywords: reverse transcriptase

Query Match 33.3%; Score 57; DB 1; Length 1281;
Best Local Similarity 41.7%; Pred. No. 19;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 10 KEVOELNSALHQSGLIDIYRTLHP 33
:::|::|::|::|::|::|
Db 170 RDTVKLTVEVKKQMOLDIYRTFYFP 193

RESULT 10
hypothetical protein yfjB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: F69805
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler,
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.,
Koetter, P.; Koningsstein, G.; Krogh, K.; Kumano, M.; Kurita, K.; Lapides, A.; Lardinois,
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A.; Authors: Schleif, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrstra, P.; Togononi, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yanane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.
A.; Authors: Yoshikawa, H.F.; Zumstede, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F69805
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A;Cross-references: UNIPROT:I031557; GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CAB12645.1
A;Experimental source: strain 168
C;Genetics:
A;Gene: yfjB
C;Superfamily: Bacillus subtilis hypothetical protein yfjB

Query Match 32.2%; Score 55; DB 2; Length 407;
Best Local Similarity 36.7%; Pred. No. 9;9;
Matches 11; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Qy 2 GTGAPRFTRKEVOLNSALHQSGLIDIYRTL 31
:::|::|::|::|::|::|
Db 118 GLGLTKFKVPQSGULTAQHCKLLLDISRAV 147

RESULT 11
S18542
hypothetical protein 4 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S18542
R;Caballero, J.L.S.; Martinez, E.; Malpartida, F.; Hopwood, D.A.

Mol. Gen. Genet. 230, 401-412, 1991
 A:Title: Organisation and functions of the actva region of the actinorhodin biosynthetic
 A:Reference number: S18539; MUID:92114870; PMID:1766437
 A:Accession: S18542
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-294 <CR>
 A:Cross-references: UNIPROT:Q53906; EMBL:X58833; NID:g46812; PIDN:CAA41640.1; PID:g46816

Query Match 31.9%; Score 54.5; DB 2; Length 294;
 Best Local Similarity 52.2%; Pred. No. 8.1;
 Matches 12; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY 4 GAPRFKEVQELNSALHQSGLID 26
 DB 40 GAPK-AKELRGASLHTGDL 61

RESULT 12
 T21973
 hypothetical protein F38H4.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T21973
 R:Lennard, N.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: Z19496
 A:Accession: T21973
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-501 <WIL>
 A:Cross-references: UNIPROT:Q20179; EMBL:Z77660; PIDN:CAB01172.1; GSPDB:GN000022; CESP:F3
 A:Experimental source: clone F38H4
 C:Genetics:
 A:Gene: CESP:F38H4.4
 A:Map position: 4
 A:Introns: 15/1; 57/3; 203/3; 311/2; 349/2; 431/3; 460/3

Query Match 31.6%; Score 54; DB 2; Length 501;
 Best Local Similarity 40.0%; Pred. No. 17;
 Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 GTGAPRFKEVQELNSALHQS 21
 DB 366 GNGSPKFKVQKIQVLRK 385

RESULT 13
 S77850
 Probable histidine-tRNA ligase (EC 6.1.1.21) - Mycoplasma capricolum (fragment)
 N:Alternate names: histidyl-tRNA synthetase; protein MC191
 C:Species: Mycoplasma capricolum
 C>Date: 09-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 09-Jul-2004
 C:Accession: S77850
 R:Bork, P.; Ouzounis, C.; Casari, G.; Schneider, R.; Sander, C.; Dolan, M.; Gilbert, W.
 Mol. Microbiol. 16, 955-967, 1995
 A:Title: Exploring the Mycoplasma capricolum genome: a minimal cell reveals its physiol
 A:Reference number: S77739; MUID:96059641; PMID:7476192
 A:Accession: S77850
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-178 <BOR>
 A:Cross-references: UNIPROT:Q49027; EMBL:Z33137; NID:g516219; PIDN:CAA83774.1; PID:g5304
 A:Experimental source: ATCC 27343
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
 C:Genetics:
 A:Genetic code: SGC3
 C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 31.3%; Score 53.5; DB 2; Length 178;
 Best Local Similarity 42.9%; Pred. No. 6.3;
 Matches 12; Conservative 7; Mismatches 8; Indels 1; Gaps 1;

QY 2 GTGAPRFKEVQELNSALHQS-DLIDY 28
 DB 65 GMLERFINIIEQNNKLNKNDQSIDLY 92

RESULT 14
 E70798
 hypothetical protein rv3740c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
 C:Accession: E70798
 R:Coile, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: E70798
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-448 <COL>
 A:Cross-references: UNIPROT:O69707; GB:AL022121; GB:AL123456; NID:g3261559; PIDN:CAA1806;
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3740c

Query Match 30.7%; Score 52.5; DB 2; Length 448;
 Best Local Similarity 48.1%; Pred. No. 25;
 Matches 13; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

QY 5 APRFIKEVQELNSALHQS-DLIDYRTL 31
 DB 89 APCRVELLELTSLH-SNLLDRHPL 114

RESULT 15
 S36565
 L2 protein - human papillomavirus type 45
 C:Species: human papillomavirus type 45
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
 C:Accession: S36565
 R:Dein, H.; Hofmann, B.
 submitted to the EMBL Data Library, August 1993
 A:Description: Primer-directed sequencing of human papillomavirus types.
 A:Reference number: S36469
 A:Accession: S36565
 A:Molecule type: DNA
 A:Residues: 1-463
 A:Cross-references: UNIPROT:P36761; EMBL:X74479; NID:g397022; PIDN:CAA52577.1; PID:g39702;
 C:Superfamily: papillomavirus L2 protein
 C:Keywords: late protein

Query Match 30.4%; Score 52; DB 2; Length 463;
 Best Local Similarity 47.8%; Pred. No. 30;
 Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 11 EVQELNSALHQS-DLIDYRTLHP 33
 DB 333 ELQPLISATNDSOLFVYADPPP 355

Search completed: November 10, 2004, 12:29:18
 Job time : 8.6713 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01; Search time 36.1065 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-14

Perfect score: 171

Sequence: 1 TGTGAPRFKEVQELNSALHQSALDIDYRTLHP 33

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Uniprot 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125.5	73.4	1192	2 Q7K241	Q7K241 homo sapien
2	125.5	73.4	1259	1 LINI HUMAN	P08547 homo sapien
3	125.5	73.4	1275	2 O00360	O00360 homo sapien
4	125.5	73.4	1275	2 O00362	O00362 homo sapien
5	125.5	73.4	1275	2 O00368	O00368 homo sapien
6	125.5	73.4	1275	2 O00370	O00370 homo sapien
7	125.5	73.4	1275	2 O00375	O00375 homo sapien
8	125.5	73.4	1275	2 O00378	O00378 homo sapien
9	125.5	73.4	1275	2 O8T30	O8T30 homo sapien
10	121.5	71.1	1275	2 O00372	O00372 homo sapien
11	119.5	69.9	1275	2 O00366	O00366 homo sapien
12	118.5	69.3	314	2 O6ZNC3	O6Znc3 homo sapien
13	118.5	69.3	314	2 BAD18452	Bad18452 homo sapi
14	118.5	69.3	1275	2 O00363	O00363 homo sapien
15	117.5	68.7	202	2 O6ZND9	O6Znd9 homo sapien
16	117.5	68.7	202	2 BAD18436	Bad18436 homo sapi
17	111	64.9	177	2 O6ZNE0	O6zne0 homo sapien
18	111	64.9	177	2 BAD18435	Bad18435 homo sapi
19	105.5	61.7	133	2 O6ZRI9	O6zri9 homo sapien
20	105.5	61.7	133	2 BAC87321	Bac87321 homo sapi
21	97	56.7	246	2 O8N281	O8n281 homo sapien
22	95.5	55.8	264	2 O6ZNC2	O6znc2 homo sapien
23	95.5	55.8	264	2 BAD18453	Bad18453 homo sapi
24	82.5	48.2	244	2 O6ZN98	O6zn98 homo sapien
25	82.5	48.2	244	2 BAD18477	Bad18477 homo sapi
26	78	45.6	1260	1 LINI NYCCO	P08548 nrycticebus
27	78	45.6	1275	2 O6Z658	O6z658 canis fami
28	75.5	44.2	152	2 O9Y443	O9y443 homo sapien
29	73.5	43.0	131	2 O6ZRN5	O6zrn5 homo sapien
30	73.5	43.0	131	2 BAC87275	Bac87275 homo sapi
31	71	41.5	160	2 O6ZSV0	O6zsv0 homo sapien

ALIGNMENTS

RESULT 1

Q7K241	PRELIMINARY;	PRT; 1192 AA.
AC Q7K241		
DT 05-JUL-2004 (Tremblrel. 27, Created)		
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)		
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)		
DE Hypothetical protein (Fragment).		
OS Homo sapiens (Human).		
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX NCBI_TaxID=9606;		
RN [1]		
RP SEQUENCE FROM N.A.		
RX MEDLINE=89233117; PubMed=2497061;		
RA Woods-Samuels P.; Wong C.; Mathias S.L.; Scott A.F.,		
RA Kazarian H.H. Jr.; Antonarakis S.E.;		
RT "Characterization of a nondeleterious L1 insertion in an intron of the		
RT human factor VIII gene and further evidence of open reading frames in		
RT functional L1 elements";		
RL Genomics 4:290-296(1989).		
RN [2]		
RP SEQUENCE FROM N.A.		
RA Woods-Samuels P.;		
RL Submitted (APR-1989) to the EMBL/GenBank/DBJ databases.		
DR EMBL; M22333; AAA88037.1; -		
DR InterPro; IPR005135; Exo_endo_phos.		
DR InterPro; IPR000477; RVTse.		
DR Pfam; PF03372; Exo_endo_phos; 1.		
DR Pfam; PF00078; RVT; 1.		
KW Hypothetical protein; RNA-directed DNA polymerase; Transferase.		
FT NON_TER		
SQ SEQUENCE 1192 AA; 139677 MW; D706D841DAE0DAD4 CRC64;		

Query Match 73.4%; Score 125.5; DB 2; Length 1192;
Best Local Similarity 42.6%; Fred.No. 2.6e-08;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;
QY 3 TGAPRFKEV-----QELNSALHQSALI 25
Db 36 TGAPRFKEVLDLQRLDLSHLINGDNTPLSLDRTRQKWKDTQELNSALHQSALI 95
QY 26 DIYRTLHP 33
Db 96 DIYRTLHP 103

RESULT 2

LINI_HUMAN	STANDARD;	PRT; 1259 AA.
ID LINI HUMAN		
AC P08547;		
DT 01-AUG-1988 (Rel. 08, Created)		
DT 01-AUG-1988 (Rel. 08, Last sequence update)		
DT 05-JUL-2004 (Rel. 44, Last annotation update)		

```

DE LINE-1 reverse transcriptase homolog.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=86230917; PubMed=2423883;
RA Hattori M., Kuhara S., Takenaka O., Sakaki Y.;
RT "Li family of repetitive DNA sequences in primates may be derived from
RL Nature 321:625-628(1986).
CC -!- MISCELLANEOUS: This sequence was constructed from an alignment of
CC published and unpublished sequences, determined in various
CC laboratories, belonging to the LINE-1 family.
DR PIR; A25913; GNHUL1.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
DR Direct protein sequencing; RNA-directed DNA polymerase.
KW SEQUENCE 1259 AA; 147112 MW; F3HC18A3803919F5 CRC64;
SQ
Query Match 73.4%; Score 125.5; DB 1; Length 1259;
Best Local Similarity 42.6%; Pred. No. 2.8e-08;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFKEV-----QELNSALHQSGLI 25
DB 118 TGAPRFKEV-----QELNSALHQSGLI 177
QY 26 DIYRTLHP 33
DB 178 DIYRTLHP 185

RESULT 3
O00360 PRELIMINARY; PRT; 1275 AA.
AC O00360;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sasanian D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93563; AAC51261.1; -.
DR PIR; B28096; B28096.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
DR RNA-directed DNA polymerase; Transferase.
KW SEQUENCE 1275 AA; 149062 MW; 350B4F0358E525F0 CRC64;
SQ
Query Match 73.4%; Score 125.5; DB 2; Length 1275;
Best Local Similarity 42.6%; Pred. No. 2.8e-08;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFKEV-----QELNSALHQSGLI 25
DB 118 TGAPRFKEV-----QELNSALHQSGLI 177
QY 26 DIYRTLHP 33
DB 178 DIYRTLHP 185

RESULT 4
O00362 PRELIMINARY; PRT; 1275 AA.
AC O00362;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sasanian D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93564; AAC51263.1; -.
DR PIR; B28096; B28096.
DR PIR; S23650; S23650.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
DR RNA-directed DNA polymerase; Transferase.
KW SEQUENCE 1275 AA; 149201 MW; 23D516D6E4358F28 CRC64;
SQ
Query Match 73.4%; Score 125.5; DB 2; Length 1275;
Best Local Similarity 42.6%; Pred. No. 2.8e-08;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFKEV-----QELNSALHQSGLI 25
DB 118 TGAPRFKEV-----QELNSALHQSGLI 178
QY 26 DIYRTLHP 33
DB 179 DIYRTLHP 186

RESULT 5
O00368 PRELIMINARY; PRT; 1275 AA.
AC O00368;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sasanian D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93568; AAC51269.1; -.

```



```
DR PIR; B28096; B28096.
DR PIR; S23650; S23650.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149077 MW; 3BBC3B2DC2E06B61 CRC64;

Query Match 73.4%; Score 125.5; DB 2; Length 1275;
Best Local Similarity 42.6%; Pred. No. 2.8e-08;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFIKEV-----QELNSALHQSGLI 25
DB 119 TGAPRFIKQVLSLDLQRLDSHTLNGDFNPLSLDRSTRQKVNKDTQELNSALHQSGLI 178
QY 26 DIYRTLHP 33
DB 179 DIYRTLHP 186

RESULT 6
ID O00370 PRELIMINARY; PRT; 1275 AA.
AC O00370;
AT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93569; AAC51271.1; -.
DR PIR; B28096; B28096.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149054 MW; 456DD0F3DD7A17F5 CRC64;

Query Match 73.4%; Score 125.5; DB 2; Length 1275;
Best Local Similarity 42.6%; Pred. No. 2.8e-08;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFIKEV-----QELNSALHQSGLI 25
DB 119 TGAPRFIKQVLSLDLQRLDSHTLNGDFNPLSLDRSTRQKVNKDTQELNSALHQSGLI 178
QY 26 DIYRTLHP 33
DB 179 DIYRTLHP 186

RESULT 8
ID O00378 PRELIMINARY; PRT; 1275 AA.
AC O00378;
AT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93574; AAC51279.1; -.
DR PIR; B28096; B28096.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149010 MW; 588703688E7129FF CRC64;

Query Match 73.4%; Score 125.5; DB 2; Length 1275;
Best Local Similarity 42.6%; Pred. No. 2.8e-08;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFIKEV-----QELNSALHQSGLI 25
DB 119 TGAPRFIKQVLSLDLQRLDSHTLNGDFNPLSLDRSTRQKVNKDTQELNSALHQSGLI 178
QY 26 DIYRTLHP 33
DB 179 DIYRTLHP 186

RESULT 7
ID O00375 PRELIMINARY; PRT; 1275 AA.
AC O00375;
AT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93572; AAC51276.1; -.
DR PIR; B28096; B28096.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149054 MW; 456DD0F3DD7A17F5 CRC64;
```

SQ SEQUENCE 1275 AA; 148879 MW; F006F74BBB72B87 CRC64;

Query Match 73.4%; Score 125.5; DB 2; Length 1275;
Best Local Similarity 42.6%; Pred. No. 2.8e-08;
Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFKEV-----QELNSALHQSGLI 25
DB 119 TGAPRFKEVLSQLRDLSDHSLTLMGDFNPLSTLDRSTRQKVNKDTQELNSALHQSGLI 178
|||||

QY 26 DIYRTLHP 33

DB 179 DIYRTLHP 186

RESULT 9

ID Q8TE30 PRELIMINARY; PRT; 1275 AA.

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21668186; PubMed=11810275;

RA Benjes S.M., Morris C.M.;

RT "A full-length and potentially active LINE element is integrated

RT polymorphically within the IGL locus in a genomically unstable region

RT of chromosome 22.";

RL Hum. Genet. 109:628-637(2001).

DR EMBL; AF421375; AAL50637.1; -.

DR PIR; B28096; B28096.

DR PIR; S23650; S23650.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.

DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.

DR InterPro; IPR005135; Exo_endo_phos.

DR InterPro; IPR000477; RVTse.

DR Pfam; PF03372; Exo_endo_phos; 1.

DR Pfam; PF00078; RVT; 1.

KW Hypothetical protein; RNA-directed DNA polymerase; Transferase.

SQ SEQUENCE 1275 AA; 149009 MW; B327D9D50A581764 CRC64;

Query Match

Best Local Similarity 73.4%; Score 125.5; DB 2; Length 1275;

Matches 29; Conservative 2; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFKEV-----QELNSALHQSGLI 25
DB 119 TGAPRFKEVLSQLRDLSDHSLTLMGDFNPLSTLDRSTRQKVNKDTQELNSALHQSGLI 178
|||||

QY 26 DIYRTLHP 33

DB 179 DIYRTLHP 186

RESULT 10

O00372

ID Q00372 PRELIMINARY; PRT; 1275 AA.

AC Q00372;

DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative p150.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sasaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
DeBerardinis R.J., Gabriel A., Swergold G.D., Kazanian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93570; AAC51273.1; -.
DR PIR; B28096; B28096.
DR PIR; S23650; S23650.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149110 MW; 4711B3BC22F7674E CRC64;

Query Match 71.1%; Score 121.5; DB 2; Length 1275;

Best Local Similarity 41.2%; Pred. No. 1e-07;

Matches 28; Conservative 2; Mismatches 1; Indels 37; Gaps 1;

QY 3 TGAPRFKEV-----KEYQELNSALHQSGLI 25
DB 119 TGAPRFKEVLSQLRDLSDHSLTLMGDFNPLSTLDRSTRQKVNKDTQELNSALHQSGLI 178
|||||

QY 26 DIYRTLHP 33

DB 179 DIYRTLHP 186

RESULT 11

O00366

ID O00366 PRELIMINARY; PRT; 1275 AA.

AC O00366;

DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative p150.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97285120; PubMed=9140393;

RA Sasaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,

DeBerardinis R.J., Gabriel A., Swergold G.D., Kazanian H.H. Jr.;

RT "Many human LI elements are capable of retrotransposition.";

RL Nat. Genet. 16:37-43(1997).

DR EMBL; U93567; AAC51267.1; -.

DR PIR; B28096; B28096.

DR PIR; S23650; S23650.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.

DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.

DR InterPro; IPR005135; Exo_endo_phos.

DR InterPro; IPR000477; RVTse.

DR Pfam; PF03372; Exo_endo_phos; 1.

DR Pfam; PF00078; RVT; 1.

KW RNA-directed DNA polymerase; Transferase.

SQ SEQUENCE 1275 AA; 149033 MW; 07B88F8F4DB831A2 CRC64;

Query Match

Best Local Similarity 69.9%; Score 119.5; DB 2; Length 1275;

Matches 28; Conservative 3; Mismatches 0; Indels 37; Gaps 1;

QY 3 TGAPRFKEV-----QELNSALHQSGLI 25
DB 119 TGAPRFKEVLSQLRDLSDHSLTLMGDFNPLSTLDRSTRQKVNKDTQELNSALHQSGLI 178
|||||

```

QY 26 DIYRTLHP 33
Db 179 DIYRTLHP 186

RESULT 12
Q6ZNC3
ID Q6ZNC3 PRELIMINARY; PRT; 314 AA.
AC Q6ZNC3;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein FLJ16220.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tongue;
RA Kanehori K., Ishibashi T., Chiba Y., Fujimori K., Hiraoka S.,
RA Tanai H., Watanabe S., Ishida S., Ono Y., Hotuta T., Watanabe M.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131275; BAD18452.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
FT NON_TER 314
SQ SEQUENCE 314 AA; 36311 MW; 2C37DD7DD1001A56 CRC64;

Query Match 69.3%; Score 118.5; DB 2; Length 314;
Best Local Similarity 39.7%; Pred. No. 5.7e-08;
Matches 27; Conservative 3; Mismatches 1; Indels 37; Gaps 1;

QY 3 TGAPRFKIVLSDVQRLDLSHTLMDGNTPLSLDRSTQKVNKHTEFNSALHQADLI 25
Db 119 TGAPRFKIVLSDVQRLDLSHTLMDGNTPLSLDRSTQKVNKHTEFNSALHQADLI 178

QY 26 DIYRTLHP 33
Db 179 DIYRTLHP 186

RESULT 13
BAD18452
ID BAD18452 PRELIMINARY; PRT; 314 AA.
AC BAD18452;
DT 12-MAY-2004 (TREMBlrel. 27, Created)
DT 12-MAY-2004 (TREMBlrel. 27, Last sequence update)
DT 12-MAY-2004 (TREMBlrel. 27, Last annotation update)
DE CDNA FLJ16220 f1s, clone CTONG3002552. (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tongue;
RA Kanehori K., Ishibashi T., Chiba Y., Fujimori K., Hiraoka S.,
RA Tanai H., Watanabe S., Ishida S., Ono Y., Hotuta T., Watanabe M.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131275; BAD18452.1;
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
FT NON_TER 314
SQ SEQUENCE 314 AA; 36311 MW; 2C37DD7DD1001A56 CRC64;

Query Match 69.3%; Score 118.5; DB 2; Length 314;
Best Local Similarity 39.7%; Pred. No. 5.7e-08;
Matches 27; Conservative 3; Mismatches 1; Indels 37; Gaps 1;

QY 3 TGAPRFKIVLSDVQRLDLSHTLMDGNTPLSLDRSTQKVNKHTEFNSALHQADLI 25
Db 119 TGAPRFKIVLSDVQRLDLSHTLMDGNTPLSLDRSTQKVNKHTEFNSALHQADLI 178

QY 26 DIYRTLHP 33
Db 179 DIYRTLHP 186

RESULT 14
O00363
ID O00363 PRELIMINARY; PRT; 1275 AA.
AC O00363;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sasaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA Deserardinis R.J., Gabriel A., Swergold G.D., Kazazian H.H. Jr.;
RT "Many human LI elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93565; AAC51264.1;
DR PIR; B28096; B28096.
DR PIR; S23650; S23650.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006278; F:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVT5e.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT; 1.
DR RNA-directed DNA polymerase; Transferase.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149007 MW; A865976EA3FD8F74 CRC64;

Query Match 69.3%; Score 118.5; DB 2; Length 1275;
Best Local Similarity 41.2%; Pred. No. 2.7e-07;
Matches 28; Conservative 2; Mismatches 1; Indels 37; Gaps 1;

QY 3 TGAPRFKIVLSDVQRLDLSHTLMDGNTPLSLDRSTQKVNKHTEFNSALHQADLI 25
Db 119 TGAPRFKIVLSDVQRLDLSHTLMDGNTPLSLDRSTQKVNKHTEFNSALHQADLI 178

QY 26 DIYRTLHP 33
Db 179 DIYRTLHP 186

RESULT 15
Q6ZND9
ID Q6ZND9 PRELIMINARY; PRT; 202 AA.
AC Q6ZND9;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein FLJ16190.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE=Thalamus;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsura N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK131258; BAD18436.1; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1
SQ SEQUENCE 202 AA; 22817 MW; B0584E72DF237584 CRC64;

Query Match 68.7%; Score 117.5; DB 2; Length 202;
Best Local Similarity 40.3%; Pred.No. 4.9e-08;
Matches 27; Conservative 3; Mismatches 0; Indels 37; Gaps 1;

Qy 4 GAPRPIKEV-----QELNSALHQSDLLD 26
Db 120 GAPRPIKQVLSDLSDLSHTIMGDNFTPLSLDRSTROKVNKDQELNLTALHQADLLD 179

Qy 27 IYRTLHP 33
Db 180 IYRTLHP 186

Search completed: November 10, 2004, 12:27:09
Job time : 37.1065 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 5.95679 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-15

Perfect score: 96

Sequence: 1 SNELTRAVELHKLKEARE 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCUTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	92.7	2349	4	US-09-538-092-914
2	48	50.0	138	3	US-09-134-001C-3167
3	48	50.0	489	4	US-09-710-279-1302
4	48	50.0	522	3	US-08-134-001C-4349
5	48	50.0	3457	2	US-08-416-603-4
6	47	49.0	528	3	US-08-987-691A-4
7	47	49.0	563	3	US-08-987-691A-2
8	47	49.0	787	4	US-09-538-092-1148
9	46	47.9	508	4	US-09-489-039A-7887
10	45	46.9	131	4	US-09-107-532A-6367
11	45	46.9	174	4	US-08-352-991A-21438
12	45	46.9	252	4	US-09-543-681A-7699
13	45	46.9	325	4	US-09-468-253B-16
14	45	46.9	341	4	US-09-468-253B-14
15	45	46.9	634	4	US-10-164-595-69
16	45	46.9	725	4	US-10-164-595-30
17	45	46.9	951	4	US-09-538-092-1231
18	44	45.8	209	4	US-09-270-767-42693
19	44	45.8	261	4	US-09-270-767-48037
20	44	45.8	340	4	US-09-468-253B-60
21	44	45.8	803	2	US-08-907-166-4
22	44	45.8	803	4	US-09-391-340-4
23	44	45.8	1333	3	US-09-356-952-2
24	44	45.8	1333	4	US-09-376-594-312
25	44	45.8	2293	3	US-09-368-590-2
26	43	44.8	478	2	US-08-951-148-1
27	43	44.8	478	2	US-09-165-234-1

28 43 44.8 478 3 US-09-274-570-1 Sequence 1, Appli
29 43 44.8 478 4 US-09-440-936-4 Sequence 4, Appli
30 43 44.8 478 4 US-09-538-092-830 Sequence 4, Appli
31 43 44.8 577 3 US-09-315-794-32 Sequence 830, Appl
32 43 44.8 577 3 US-09-389-341-32 Sequence 32, Appl
33 43 44.8 577 4 US-09-538-092-369 Sequence 32, Appl
34 42.5 44.3 822 4 US-09-489-039A-8709 Sequence 369, Appl
35 42 43.8 101 4 US-09-270-767-58656 Sequence 8709, Appl
36 42 43.8 123 4 US-09-107-532A-4652 Sequence 58656, A
37 42 43.8 252 4 US-09-270-767-43310 Sequence 4652, Ap
38 42 43.8 281 4 US-09-270-767-62321 Sequence 43310, A
39 42 43.8 336 4 US-09-468-253B-59 Sequence 62321, A
40 42 43.8 395 4 US-09-248-796A-17269 Sequence 59, Appl
41 42 43.8 492 4 US-09-270-767-46706 Sequence 17269, A
42 42 43.8 573 4 US-09-252-991A-20056 Sequence 46706, A
43 42 43.8 600 6 5240706-1 Sequence 20056, A
44 42 43.8 2101 1 US-08-466-390-4 Patent No. 5240706
45 42 43.8 2101 1 US-08-470-950-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-538-092-914
; Sequence 914, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 914
; LENGTH: 2349
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P12270
US-09-538-092-914

Query Match 92.7%; Score 89; DB 4; Length 2349;
Best Local Similarity 95.0%; Pred. No. 0.00086;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SNELTRAVELHKLKEARE 20
DB 300 SNELTRAVELHKLKEAGE 319
|||||

RESULT 2
US-09-134-001C-3167
; Sequence 3167, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/987,691A
FILING DATE: 09-NOV-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/047,955
FILING DATE: 29-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Braman, Susan J
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 001.00151
TELEPHONE: 716-393-3002
TELEFAX: 716-393-3001
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-987-691A-4

Query Match 49.0%; Score 47; DB 3; Length 528;
Best Local Similarity 42.1%; Pred. No. 87;
Matches 8; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKEAR 19
DB 263 THELTQAVKIHVIKELK 281

RESULT 7
US-08-987-691A-2
Sequence 2, Application US/08987691A
Patent No. 6306387
GENERAL INFORMATION:
APPLICANT: Galan, Jorge E
TITLE OF INVENTION: ANTIGEN DELIVERY SYSTEM
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Braman & Rogalskyj, LLP
STREET: P.O. Box 352
CITY: Canandaigua
STATE: New York
COUNTRY: USA
ZIP: 14424-0352
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/987,691A
FILING DATE: 09-NOV-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/047,955
FILING DATE: 29-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Braman, Susan J
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 001.00151
TELEPHONE: 716-393-3002
TELEFAX: 716-393-3001
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 563 amino acids
TYPE: amino acid

STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-987-691A-2

Query Match 49.0%; Score 47; DB 3; Length 563;
Best Local Similarity 42.1%; Pred. No. 92;
Matches 8; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKEAR 19
DB 263 THELTQAVKIHVIKELK 281

RESULT 8
US-09-538-092-1148
Sequence 1148, Application US/09538092
Patent No. 6753314
GENERAL INFORMATION:
APPLICANT: Giot, Loic
APPLICANT: Mansfield, Traci A.
TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
FILE REFERENCE: 15966-542
CURRENT APPLICATION NUMBER: US/09/538,092
CURRENT FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/127,352
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: 60/178,965
PRIOR FILING DATE: 2000-02-01
NUMBER OF SEQ ID NOS: 1387
SOFTWARE: CurapatSeqFormatter Version 0.9
SEQ ID NO 1148
LENGTH: 787
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: misc.feature
LOCATION: (0)...(0)
OTHER INFORMATION: Polypeptide Accession Number P49321
US-09-538-092-1148

Query Match 49.0%; Score 47; DB 4; Length 787;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKEARE 20
DB 631 SSEYKKEIEELKELLPEIRE 650

RESULT 9
US-09-489-039A-7887
Sequence 7887, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 7887
LENGTH: 508
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7887

Query Match 47.9%; Score 46; DB 4; Length 508;
Best Local Similarity 45.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKREAR 20
DB 264 TRELTRLVSNLRLVRSERE 283

RESULT 10

US-09-107-532A-6367
; Sequence 6367, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/885,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6367:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...131
; SEQUENCE DESCRIPTION: SEQ ID NO: 6367:
US-09-107-532A-6367

Query Match 46.9%; Score 45; DB 4; Length 131;
Best Local Similarity 66.7%; Pred. No. 44;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 ELTRAVEELHKLKKE 17
DB 86 EKTRAVSELTQLMKE 100

RESULT 11

US-09-252-991A-21438
; Sequence 21438, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21438
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21438

Query Match 46.9%; Score 45; DB 4; Length 174;
Best Local Similarity 55.6%; Pred. No. 57;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 NELTRAVSELHKLKREAR 19
DB 46 NELQSAABELNAMLQYAR 63

RESULT 12

US-09-543-681A-7699
; Sequence 7699, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7699
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7699

Query Match 46.9%; Score 45; DB 4; Length 252;
Best Local Similarity 56.2%; Pred. No. 81;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LTRAVEELHKLKREAR 19
DB 236 LAPAVEEMKRLMREAR 251

RESULT 13

US-09-468-253B-16
; Sequence 16, Application US/09468253B
; Patent No. 6750012
; GENERAL INFORMATION:
; APPLICANT: Gould-Rothberg, Bonnie
; TITLE OF INVENTION: Method of Identifying a Psychotropic Agent Using
; TITLE OF INVENTION: Differential Gene Expression
; FILE REFERENCE: 15966-519
; CURRENT APPLICATION NUMBER: US/09/468,253B
; CURRENT FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: 60/113,127
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-468-253B-16

Query Match 46.9%; Score 45; DB 4; Length 325;
Best Local Similarity 61.1%; Pred. No. 1.e+02;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 ELTRAVEELHKLKKEARE 20
:|||||:|||||
Db 167 KLKEAVEEVKKLLVPAAE 184

RESULT 14

US-09-468-253B-14
; Sequence 14, Application US/09468253B
; Patent No. 6750012
; GENERAL INFORMATION:
; APPLICANT: Gould-Rothberg, Bonnie
; TITLE OF INVENTION: Method of Identifying a Psychotropic Agent Using
; TITLE OF INVENTION: Differential Gene Expression
; FILE REFERENCE: 15966-519
; CURRENT APPLICATION NUMBER: US/09/468,253B
; CURRENT FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: 60/113,127
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-468-253B-14

Query Match 46.9%; Score 45; DB 4; Length 341;
Best Local Similarity 61.1%; Pred. No. 1.e+02;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 ELTRAVEELHKLKKEARE 20
:|||||:|||||
Db 167 KLKEAVEEVKKLLVPAAE 184

RESULT 15

US-10-164-595-69
; Sequence 69, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: IU 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE: 2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69
; LENGTH: 634
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-595-69

Query Match 46.9%; Score 45; DB 4; Length 634;
Best Local Similarity 56.2%; Pred. No. 1.9e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 SNELTRAVEELHKLK 16
:|||||:|||||
Db 338 SNAAIRKVKVKKHLE 353

Search completed: November 10, 2004, 12:32:28
Job time : 6.95679 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 15:53:52 ; Search time 19,2901 Seconds
(without alignments)
366.225 Million cell updates/sec

Title: US-10-092-750-15
Perfect score: 96
Sequence: 1 SNELTRAVEELHKLKEARE 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA.*
- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/prodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	20	14	US-10-092-750-15
2	89	92.7	122	14	Sequence 15, Appl
3	53	55.2	452	16	Sequence 244, App
4	51	53.1	418	16	Sequence 136019
5	50	52.1	339	15	Sequence 163252
6	50	52.1	339	15	Sequence 160411
7	49	51.0	87	15	Sequence 160467
8	49	51.0	294	17	Sequence 186169
9	49	51.0	314	16	Sequence 8785, Ap
10	49	51.0	328	16	Sequence 135978
11	49	51.0	330	15	Sequence 135893
12	49	51.0	330	15	Sequence 160433
13	49	51.0	339	15	Sequence 160478
					Sequence 257223

14 49 51.0 392 16 US-10-437-963-170935 Sequence 170935,
15 49 51.0 483 16 US-10-437-963-135858 Sequence 135858,
16 49 51.0 545 16 US-10-437-963-170941 Sequence 170941,
17 49 51.0 640 16 US-10-437-963-135898 Sequence 135898,
18 49 51.0 659 16 US-10-437-963-135852 Sequence 135852,
19 49 51.0 679 16 US-10-437-963-135990 Sequence 135990,
20 49 51.0 730 16 US-10-437-963-135903 Sequence 135903,
21 49 51.0 799 16 US-10-437-963-135946 Sequence 135946,
22 49 51.0 810 16 US-10-437-963-135985 Sequence 135985,
23 49 51.0 996 16 US-10-437-963-135853 Sequence 135853,
24 49 51.0 1138 16 US-10-437-963-136023 Sequence 136023,
25 49 51.0 1159 16 US-10-437-963-135940 Sequence 135940,
26 49 51.0 1222 16 US-10-437-963-135849 Sequence 135849,
27 49 51.0 1326 16 US-10-437-963-135935 Sequence 135935,
28 49 51.0 2048 16 US-10-437-963-183905 Sequence 183905,
29 48 50.0 112 15 US-10-282-122A-70584 Sequence 70584, A
30 48 50.0 114 15 US-10-282-122A-70340 Sequence 70340, A
31 48 50.0 130 9 US-09-815-242-12607 Sequence 12607, A
32 48 50.0 130 9 US-09-815-242-12750 Sequence 12750, A
33 48 50.0 489 15 US-10-282-122A-71002 Sequence 71002, A
34 48 50.0 973 16 US-10-437-963-135992 Sequence 135992,
35 48 50.0 1392 13 US-10-087-192-252 Sequence 252, App
36 48 50.0 2859 13 US-10-087-192-249 Sequence 249, App
37 47 49.0 27 9 US-09-739-907-165 Sequence 165, App
38 47 49.0 27 11 US-09-938-671-165 Sequence 165, App
39 47 49.0 121 15 US-10-424-599-157958 Sequence 157958,
40 47 49.0 123 9 US-09-764-868-922 Sequence 922, App
41 47 49.0 125 9 US-09-764-868-1228 Sequence 1228, Ap
42 47 49.0 173 9 US-09-739-907-162 Sequence 162, App
43 47 49.0 173 11 US-09-938-671-162 Sequence 162, App
44 47 49.0 267 14 US-10-104-047-2252 Sequence 2252, Ap
45 47 49.0 543 17 US-10-478-516-13 Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-10-092-750-15
; Sequence 15, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-XL
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-15

Query Match 100.0%; Score 96; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKEARE 20
Db 1 SNELTRAVEELHKLKEARE 20

RESULT 2
US-10-092-750-244
; Sequence 244, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:

```
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 5036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-244

Query Match          92.7%; Score 89; DB 14; Length 122;
Best Local Similarity 95.0%; Pred. No. 8.2e-05;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKKEARE 20
   |||||
Db 31 SNELTRAVEELHKLKKEAGE 50

RESULT 3
US-10-437-963-136019
; Sequence 136019, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 136019
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37638C.1.pep
US-10-437-963-136019

Query Match          55.2%; Score 53; DB 16; Length 452;
Best Local Similarity 52.6%; Pred. No. 34;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 NELTRAVEELHKLKKEARE 20
   |||||
Db 158 NNLNRTLAELHRLKLTAE 176

RESULT 4
US-10-437-963-163252
; Sequence 163252, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
US-10-437-963-163252

; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 163252
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_62266C.1.pep
US-10-437-963-163252

Query Match          53.1%; Score 51; DB 16; Length 418;
Best Local Similarity 58.8%; Pred. No. 60;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 NELTRAVEELHKLKKEA 18
   |||||
Db 389 NELANHIEQLRKLUREA 405

RESULT 5
US-10-424-599-160411
; Sequence 160411, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 160411
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(329)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_11586C.1.pep
US-10-424-599-160411

Query Match          52.1%; Score 50; DB 15; Length 329;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKKEARE 20
   |||||
Db 72 AKEYEKAIEELQKLREKSE 91

RESULT 6
US-10-424-599-160467
; Sequence 160467, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
```

```
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 150467
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(339)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_11591C.1.pbp
US-10-424-599-160467

Query Match          52.1%; Score 50; DB 15; Length 339;
Best Local Similarity 50.0%; Pred. No. 66;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 SNETRAVELHKLKEARE 20
   :|||:|||||:|
Db 72 AQEYKAIEELQKLLREKSE 91

RESULT 7
US-10-424-599-186169
; Sequence 186169, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 186169
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_139122C.1.pbp
US-10-424-599-186169

Query Match          51.0%; Score 49; DB 15; Length 87;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 SNETRAVELHKLKEARE 20
   :|||:|||||:|
Db 61 AQEYKAIEELQKLLREKSE 80

RESULT 8
US-10-739-930-8785
; Sequence 8785, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 8785
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(294)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: GLYMA-23APR03-C1728_56.p
US-10-739-930-8785

Query Match          51.0%; Score 49; DB 17; Length 294;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 SNETRAVELHKLKEARE 20
   :|||:|||||:|
Db 38 AQEYKAIEELQKLLREKSE 57

RESULT 9
US-10-437-963-135978
; Sequence 135978, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 135978
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37600C.1.pbp
US-10-437-963-135978

Query Match          51.0%; Score 49; DB 16; Length 314;
Best Local Similarity 52.6%; Pred. No. 84;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 NELTRAVEELHKLKEARE 20
   :|||:|||||:|
Db 205 NNLARTLAELHGLKLTAE 223

RESULT 10
US-10-437-963-135893
; Sequence 135893, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 135893
; LENGTH: 328
```

; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37524C.1.pep
US-10-437-963-135893

Query Match 51.0%; Score 49; DB 16; Length 328;
Best Local Similarity 52.6%; Pred. No. 88;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 NELTRAVEELHKLKKEARE 20
Db 99 NNLNRTLAEHLGMLKTAEE 117

RESULT 11

US-10-424-599-160433
; Sequence 160433, Application US/10424599
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 160433
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Glycine max

FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_11588C.1.pep

US-10-424-599-160433

Query Match 51.0%; Score 49; DB 15; Length 330;
Best Local Similarity 50.0%; Pred. No. 89;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 SNETLRAVEELHKLKKEARE 20
Db 74 AKEYEQAELEQLKREKSE 93

RESULT 12

US-10-424-599-160478
; Sequence 160478, Application US/10424599
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 160478
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Glycine max

FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_11592C.1.pep

US-10-424-599-160478

Query Match 51.0%; Score 49; DB 15; Length 339;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 SNETLRAVEELHKLKKEARE 20
Db 72 AKEYEQAELEQLKREKSE 91

RESULT 13

US-10-424-599-257223
; Sequence 257223, Application US/10424599
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 257223
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Glycine max

FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_74295C.1.pep

US-10-424-599-257223

Query Match 51.0%; Score 49; DB 15; Length 345;
Best Local Similarity 40.0%; Pred. No. 93;
Matches 8; Conservative 10; Mismatches 2; Indels 0; Gaps 0;

QY 1 SNETLRAVEELHKLKKEARE 20
Db 286 SSETQSTEEILAVQEAR 305

RESULT 14

US-10-437-963-170935
; Sequence 170935, Application US/10437963
; Publication No. US20040123343A1

GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping

TITLE OF INVENTION:

Rice Nucleic Acid Molecules and Other Molecules Associated With

Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 170935

LENGTH: 392

TYPE: PRT

ORGANISM: Oryza sativa

FEATURE:

OTHER INFORMATION: Clone ID: PAT_MRT4530_69213C.1.pep

US-10-437-963-170935

Query Match 51.0%; Score 49; DB 16; Length 392;
Best Local Similarity 52.6%; Pred. No. 1.1e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 NELTRAVEELHKLKKEARE 20
Db 314 NNLNRTLAEHLGMLKTAEE 332

```
RESULT 15
US-10-437-963-135858
; Sequence 135858, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 135858
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(483)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37493C.1.pep
US-10-437-963-135858

Query Match          51.0%; Score 49; DB 16; Length 483;
Best Local Similarity 52.6%; Pred. No. 1.3e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY      2 NELTRAVEELHKLKEARE 20
Db      45 NNLNRTLAELHGMKTAEE 63

Search completed: November 10, 2004, 16:35:59
Job time : 20.2901 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:36:51 ; Search time 4.04321 Seconds
(without alignments)
475.942 Million cell updates/sec

Title: US-10-092-750-15

Perfect score: 96
Sequence: 1 SNELTRAVEELHKLKEARE 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	92.7	2094	2 S33124	tpr protein - huma
2	51	53.1	587	2 S29787	UDPgalacturonate d
3	50	52.1	156	2 C82716	ATP synthase, B ch
4	49	51.0	284	2 A40689	assemblin, striate
5	49	51.0	339	2 E71169	hypothetical prote
6	48	50.0	114	2 F89322	conserved hypothet
7	48	50.0	395	2 AC1754	capsid protein (ba
8	48	50.0	822	2 F86812	phosphoketolase (i
9	48	50.0	1849	2 T00415	hypothetical prote
10	47	49.0	418	2 T25368	hypothetical prote
11	47	49.0	543	2 AB0850	tyrosine phosphata
12	47	49.0	787	2 A48819	nuclear autoantige
13	46	47.9	275	2 D69749	transcription regu
14	46	47.9	305	2 C85042	SYR1-like syntaxin
15	46	47.9	460	2 G85686	sensor protein Pho
16	46	47.9	486	1 B41966	sensor kinase phoQ
17	46	47.9	486	2 A99829	sensor protein Pho
18	46	47.9	580	2 T36393	L-aspartate oxidas
19	46	47.9	680	2 A43800	nuclear autoantige
20	46	47.9	2763	1 WZBE22	gene 22 protein -
21	45.5	47.4	140	1 R3HS9H	ribosomal protein
22	45.5	47.4	140	2 H84321	30S ribosomal prot
23	45	46.9	140	1 R3HS19	ribosomal protein
24	45	46.9	168	2 H83075	type 4 fibrial bi
25	45	46.9	215	2 D82872	deoxyribose-phosph
26	45	46.9	251	2 B39685	GRESAG protein 2.1
27	45	46.9	403	2 T15136	hypothetical prote
28	45	46.9	419	2 T48080	translation releas
29	45	46.9	463	2 C86034	L-seryl-tRNAse se

30 45 46.9 463 2 D91187 L-seryl-tRNAse se
31 45 46.9 463 2 A65159 L-seryl-tRNAse se
32 45 46.9 487 1 ZGEBPT sensor kinase phoQ
33 45 46.9 487 2 AG0646 sensor protein Pho
34 45 46.9 677 2 T19281 hypothetical prote
35 45 46.9 885 2 AC2179 hypothetical prote
36 45 46.9 961 2 A55380 facicogenital dyspl
37 45 46.9 962 2 C81060 translation initia
38 45 46.9 962 2 A81817 translation initia
39 45 46.9 1070 2 T31069 tolloid-BMP-1 like
40 44.5 46.4 864 2 B90395 purine NTPase (imp
41 44 45.8 111 2 A70945 hypothetical prote
42 44 45.8 251 2 A86727 transcription regu
43 44 45.8 265 2 T09570 carbonate dehydrat
44 44 45.8 281 2 AD1745 B. subtilis YumP p
45 44 45.8 293 2 B70455 thiosulfate sulfur

ALIGNMENTS

RESULT 1

S33124
tpr protein - human
N:Alternate names: kinase-related transforming protein (tpr-net); protein with promoter 1
C:Species: Homo sapiens (man)
C:Date: 22-Nov-1993 #sequence.Revision 26-May-1995 #text_change 09-Jul-2004
C:Accession: S33124; S23740; S00928; G01185
R:Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 2329-2333, 1992
A:Title: The human tpr gene encodes a protein of 2094 amino acids that has extensive coil
A:Reference number: S33124; MUID:93064711; PMID:1437155
A:Accession: S33124
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Cross-references: 1-2094 <MIT>
A:Cross-references: UNIPROT:Q15624; UNIPROT:Q9UEB3; EMBL:X66397; NID:G633225
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
R:Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 383-388, 1992
A:Title: Nucleotide sequence analysis of human tpr cDNA clones.
A:Reference number: S23740; MUID:92195670; PMID:1549355
A:Accession: S23740
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-725, 'L', <MT2>
A:Cross-references: EMBL:X63105; NID:g37257; PIDN:CAA44819.1; PID:G37258
R:King, H.W.S.; Tempst, P.R.; Merrifield, K.R.; Rance, A.J.
Oncogene 2, 617-619, 1988
A:Title: Tpr homologues activate met and raf.
A:Reference number: S00928; MUID:88262257; PMID:3387099
A:Accession: S00928
A:Molecule type: mRNA
A:Residues: 1-31, 'R', 33-142 <KIN>
A:Cross-references: EMBL:Y00672; NID:g37255; PIDN:CAA68681.1; PID:g37256
R:Greco, A.
Submitted to the EMBL Data Library, December 1995
A:Reference number: H00592
A:Accession: G01185
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 144-228 <GRE>
A:Cross-references: EMBL:X94208; NID:g1296797; PIDN:CAA63904.1; PID:g1296798
C:Genetics:
A:Gene: GDB:TPR
A:Cross-references: GDB:128821; OMIM:189940
A:Map position: 1q25-1q25
A:Introns: 177/3

Query Match 92.7%; Score 89; DB 2; Length 2094;
Best Local Similarity 95.0%; Pred. No. 0.00081;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKEARE 20
|||||
Db 300 SNELTRAVEELHKLKEAGE 319

RESULT 2
S29787
UDPgalacturonate decarboxylase (EC 4.1.1.67) alpha chain - Acidaminococcus fermentans
N:Alternate names: UDPgalacturonate decarboxylase carboxytransferase chain
C:Species: Acidaminococcus fermentans
C>Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S29787; S61762; S36104; S28382
R:Bendrat, K.; Buckel, W.
Eur. J. Biochem. 211, 697-702, 1993
A:Title: Cloning, sequencing and expression of the gene encoding the carboxytransferase
erichia coli.
A:Reference number: S29787; MUID:93170302; PMID:8382157
A:Accession: S29787
A:Molecule type: DNA
A:Residues: 1-587 <BEN>
A:Cross-references: UNIPROT:Q06700; EMBL:X69435; NID:g49181; PIDN:CAA49210.1; PID:g49182
A:Accession: S61762
A:Molecule type: protein
A:Residues: 'X', 3-22, 'X', 24-35; 454-461 <BEW>
R:Bendrat, K.; Mueller, U.; Klees, A.G.; Buckel, W.
FEBS Lett. 329, 329-331, 1993
A:Title: Identification of the gene encoding the activator of (R)-2-hydroxyglutaryl-CoA
A:Reference number: S36104; MUID:93374040; PMID:8365476
A:Accession: S36104
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 548-587 <BE2>
A:Cross-references: EMBL:X59645; NID:g433930; PIDN:CAA42195.1; PID:g433931
C:Genetics:
A:Gene: gcaA
C:Superfamily: propionyl-CoA carboxylase beta chain
C:Keywords: carbon-carbon lyase; carboxy-lyase
F:2-587/Product: UDPgalacturonate decarboxylase alpha chain #status experimental <MAT>

Query Match 53.1%; Score 51; DB 2; Length 587;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 ELTRAVEELHKLKEARE 20
|||:|:|:|:|:|:|
Db 30 OLKKIEEIHQLKEAGE 47

RESULT 3
C82716
ATP synthase, B chain Xf1147 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: C82716
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20355717; PMID:10910347
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: C82716
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STM>
A:Cross-references: UNIPROT:Q9PE81; GB:AE003950; GB:AE003849; NID:g9106104; PIDN:AAF8395
R:Simpson, A.G.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
de-Neto, E.; Docena, C.; El-Dorzy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.N.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.W.; Silva Jr., W.A.; da Silveira
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
C:Contents: annotation
C:Genetics:
A:Gene: Xf1147
C:Superfamily: H⁺-transporting ATP synthase chain I

Query Match 52.1%; Score 50; DB 2; Length 156;
Best Local Similarity 61.1%; Pred. No. 11;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 ELTRAVEELHKLKEARE 20
|||:|:|:|:|:|:|
Db 53 ELAQOEIKTKLNARE 70

RESULT 4
A40689
N:Alternate names: striated fiber - green alga (Spermatopsis similis)
C:Species: Spermatopsis similis
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A40689; S29746
R:Weber, K.; Gaister, N.; Plessmann, U.; Bremerich, A.; Lechtreck, K.F.; Melkonian, M.
J. Cell Biol. 121, 837-845, 1993
A:Title: SF-assemblin, the structural protein of the 2-nm filaments from striated microti
A:Reference number: A40689; MUID:93260016; PMID:8491776
A:Accession: A40689
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-284 <WEB>
A:Cross-references: UNIPROT:P55925

Query Match 51.0%; Score 49; DB 2; Length 284;
Best Local Similarity 47.1%; Pred. No. 26;
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 4 LTRAVEELHKLKEARE 20
|||:|:|:|:|:|:|
Db 133 LSRTMQDLHTTIKEARE 149

RESULT 5
E71169
hypothetical protein PH0553 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C:Accession: E71169
R:Kawabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine
M.; Ohuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71169
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-399 <KAW>
A:Cross-references: UNIPROT:O58288; GB:AP000002; NID:g3236129; PIDN:BAA29642.1; PID:g3258
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0553

Query Match 51.0%; Score 49; DB 2; Length 399;
Best Local Similarity 47.4%; Pred. No. 37;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 NELTRAVEELHKLKEARE 20
|||:|:|:|:|:|:|
Db 134 SELAKKVEEHKKVEELE 152

RESULT 6
F89322
conserved hypothetical protein SA1279 [imported] - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-Nov-2003
C/Accession: F89322
R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogun-
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Panel 357, 1423-1230, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89922
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-114 <KUR>
A:Cross-references: GB|BA000018; PID:g13701244; PIDN:BAB42539.1; GSPDB:GN00149

A; Experimental source: strain NC12
 C; Genetics:
 A; Gene: SA1279
 C; Superfamily: uncharacterized conserved protein

 Query Match 50.0%; Score 48; DB 2; Length 114;
 Best Local Similarity 50.0%; Pred. No. 15;
 Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

 Qy 1 SNETRAVEELHKKLKEARE 20
 : : : : :
 Db 49 NNEVKLSEENHKKLKEEE 68

RESULT 7
AC1754
capsid protein [bacteriophage b1285] homolog lin2576 [imported] - *Listeria innocua* (strain 6407)
C/Species: *Listeria innocua*
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AC1754

D.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001

ok, C.; Schluter, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Titte: Comparative genomics of *Listeria* species. A:Reference number: AB1077; MUID:21537279; PMID:11673669

A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1395 <GLA>
A/Cross-references: UNIPROT:Q928F9; GB:AL592022; PIDN:CAC97803.1; PID:G16415098; GSPDB:C
A/Experimental source: strain Clp11262
C/Genetics:
A/Gene: lin2576

```

QY      2  NELTRAVEELHKLKEARE  20
      :|||: |||: |||: |||:
Db      71  DELTAKIEELEKLEEAAND  89
Query Match      50.0%; Score 48; DB 2; Length 395;
Best Local Similarity 52.6%; Pred.No. 50;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Caps 0;

```

RESULT 8
F86812
phosphoketolase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C/Accession: F86812
R/Bolotin, A.; Wincker, P.; Manger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001

C;Accession: G85686
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: G85686
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-460 <STO>
A;Cross-references: UNIPROT:Q8X739; GB:AE005174; NID:G12514780; PIDN:AAG55955.1; GSPDB:C
A;Experimental source: strain O157:H7, substrain EDL533
C;Genetics:
A;Gene: phoQ
C;Superfamily: envZ protein; sensor histidine kinase homology

Query Match 47.9%; Score 46; DB 2; Length 460;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKKEARE 20
: ||| | | : ||| | |
Db 218 TRELTSLVNRLKSERE 237

Search completed: November 10, 2004, 12:29:19
Job time : 5.04321 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01 ; Search time 21.8827 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-15

Perfect score: 96

Sequence: 1 SNELTRAVEELHKLKEARE 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	89	92.7	726	Q15624	Q15624 homo sapien
2	89	92.7	2349	TPR_HUMAN	P12270 homo sapien
3	89	92.7	2363	Q99568	Q99568 homo sapien
4	86	89.6	411	Q8BK71	Q8BK71 mus musculus
5	86	89.6	681	Q8BU18	Q8BU18 mus musculus
6	86	89.6	1200	Q921B9	Q921B9 mus musculus
7	86	89.6	2357	Q7M739	Q7M739 mus musculus
8	51	53.1	163	Q7P5L1	Q7P5L1 fuscobacteri
9	51	53.1	392	Q732Z3	Q732Z3 mycobacteri
10	51	53.1	392	QAS03775	QAS03775 mycobacte
11	51	53.1	418	Q84YR9	Q84YR9 oryza sativ
12	51	53.1	587	Q6DA_ACIFE	Q6DA_ACIFE
13	50	52.1	156	Q9P831	Q9P831 xylella fas
14	50	52.1	648	Q7UWL4	Q7UWL4 rhodopirell
15	49	51.0	112	RPOL_METKA	Q8T281 methanopyru
16	49	51.0	284	SPAS_SPESI	P55925 spermatozop
17	49	51.0	399	Q58288	Q58288 pyrococcus
18	49	51.0	948	Q7XX19	Q7XX19 oryza sativ
19	49	51.0	1063	Q75H25	Q75H25 oryza sativ
20	49	51.0	1063	QAR06328	QAR06328 oryza sat
21	49	51.0	1084	Q75L70	Q75L70 oryza sativ
22	49	51.0	1084	QAS01945	QAS01945 oryza sat
23	49	51.0	1161	Q7YIM7	Q7YIM7 oryza sativ
24	49	51.0	5085	Q70LM4	Q70LM4 bacillus br
25	49	51.0	5085	CAD92852	CAD92852 bacillus
26	48	50.0	112	Q8CSJ3	Q8CSJ3 staphylococ
27	48	50.0	114	Q6G9B7	Q6G9B7 staphylococ
28	48	50.0	114	Q6GGW4	Q6GGW4 staphylococ
29	48	50.0	114	Q99U42	Q99U42 staphylococ
30	48	50.0	114	Q7A0V6	Q7A0V6 staphylococ
31	48	50.0	114	Q7A5L1	Q7A5L1 staphylococ

```

32 48 50.0 163 2 Q8RGD8
33 48 50.0 167 2 Q877H8
34 48 50.0 231 2 Q858D1
35 48 50.0 395 2 Q828F9
36 48 50.0 438 2 Q73B78
37 48 50.0 438 2 Q81T51
38 48 50.0 438 2 AAS40471
39 48 50.0 438 2 AAT30537
40 48 50.0 451 2 Q7QTR3
41 48 50.0 478 2 Q84SQ3
42 48 50.0 489 2 Q8CMK1
43 48 50.0 629 2 Q6CNW6
44 48 50.0 822 1 PHK_LACLA
45 48 50.0 943 2 Q7QD47

```

ALIGNMENTS

```

RESULT 1
Q15624
ID Q15624 PRELIMINARY; PRT; 726 AA.
AC Q15624;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Tpr protein.
OS Name=Tpr;
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92195670; PubMed=1549355;
RA Mitchell P.J., Cooper C.;
RT "Nucleotide sequence analysis of human tpr cDNA clones.";
RL Oncogene 7:383-388(1992).
DR EMBL; X63105; CAA44819.1; -.
DR PIR; S33124; S33124.
SQ SEQUENCE 726 AA; 83978 MW; E6351C8C59B6C67 CRC64;

```

```

Query Match 92.7%; Score 89; DB 2; Length 726;
Best Local Similarity 95.0%; Pred. No. 0.0033;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 SNELTRAVEELHKLKEARE 20
   |||||
DB 300 SNELTRAVEELHKLKEARE 319

```

```

RESULT 2
TPR_HUMAN
ID TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Nucleoprotein TPR.
GN Name=TPR;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
   extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.

```

```

RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "tpr", a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex."
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
RN SEQUENCE OF 1-142 FROM N.A.
RP MEDLINE=68262257; PubMed=3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RA "tpr" homologues activate met and raf."
RL Oncogene 2:617-619(1988).
CC CC -!- FUNCTION: Component of the cytoplasmic fibrils of the nuclear pore
CC complex implicated in nuclear protein import. Its amino terminus
CC is involved in activation of oncogenic kinases.
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic surface of the nuclear pore
CC complex. The assembly of the NPC is a stepwise process in which
CC tpr-containing peripheral structures assemble after other
CC components, including p62.
CC CC -!- TISSUE SPECIFICITY: Highest in testis, lung, thymus, spleen and
CC brain, lower levels in heart, liver and kidney.
CC CC -!- DISEASE: Involved in tumorigenic rearrangements with the MET, TRK
CC or RAF genes.
CC CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/TPRID282.html".
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; X66397; CAA47021.1; --
CC DR EMBL; Y00672; CAA68681.1; --
CC DR Genew; HGNC:12017; TPR.
CC DR MIM; 189940;
CC DR GO; GO:0005737; C:cytoplasm; TAS.
CC DR GO; GO:0005643; C:nuclear pore; TAS.
CC DR GO; GO:0006006; P:protein-nucleus import; TAS.
CC KW Chromosomal translocation; Coiled coil; Nuclear protein;
CC KW Proto-oncogene; Transport.
CC FT DOMAIN 78 360 Coiled coil (Potential).
CC FT DOMAIN 422 571 Coiled coil (Potential).
CC FT DOMAIN 575 828 Coiled coil (Potential).
CC FT DOMAIN 758 805 Coiled coil (Potential).
CC FT DOMAIN 834 869 Coiled coil (Potential).
CC FT DOMAIN 934 979 Coiled coil (Potential).
CC FT DOMAIN 1004 1064 Coiled coil (Potential).
CC FT DOMAIN 1138 1166 Coiled coil (Potential).
CC FT DOMAIN 1196 1241 Coiled coil (Potential).
CC FT DOMAIN 1262 1304 Coiled coil (Potential).
CC FT DOMAIN 1354 1434 Coiled coil (Potential).
CC FT DOMAIN 1476 1595 Coiled coil (Potential).
CC FT DOMAIN 527 530 Poly-Ser.
CC FT DOMAIN 1833 1836 Poly-Glu.
CC FT DOMAIN 1957 1964 Poly-Asp.
CC FT DOMAIN 2295 2298 Poly-Ser.
CC SEQUENCE 2349 AA; 285600 MW; AFD6885CEDCA9EF CRC64;
Query Match 92.7%; Score 89; DB 1; Length 2349;
Best Local Similarity 95.0%; Pred. No. 0.0095;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKKEARE 20
DB 300 SNELTRAVEELHKLKKEAGE 319

RESULT 3
Q99968

```

```

ID Q99968 PRELIMINARY; PRT; 2363 AA.
AC Q99968;
DT 01-MAY-1997 (TREMELrel. 03, Created)
DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Tpr.
GN Name=tpr;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97177132; PubMed=9024684;
RA Cordes V.C., Reidenbach S., Rackwitz H.R., Franke W.W.;
RT "Identification of protein p270/Tpr as a constitutive component of the
RT nuclear pore complex-attached intranuclear filaments."
RL J. Cell Biol. 136:515-529(1997).
DR EMBL; U69688; AA848030.1; --
SQ SEQUENCE 2363 AA; 267333 MW; E9BA1C6578AA35B0 CRC64;
Query Match 92.7%; Score 89; DB 2; Length 2363;
Best Local Similarity 95.0%; Pred. No. 0.0095;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKKEARE 20
DB 300 SNELTRAVEELHKLKKEAGE 319

RESULT 4
Q98K71
ID Q98K71 PRELIMINARY; PRT; 411 AA.
AC Q98K71;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
DE DE enriched library, clone,2610029M07 product,NUCLEAR PORE COMPLEX-
DE ASSOCIATED PROTEIN TPR homolog (Fragment).
GN Name=Tpr;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA The FANTOM Consortium,
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Sugahara Y., Hayatsu N., Itoh M.,

```


GN Name=Tpr;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/Ola;
 RA Kuznetsov N.V., Sandblad L., Hase M.E., Hunziker A., Hergt M.,
 RA Cordes V.C.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AJ298076; CAC40701.1; -.
 DR MGD; MGI:1922066; Tpr.
 DR GO; GO:0005635; C:nuclear membrane; IDA.
 DR NON TER 1200 1200
 FT
 SQ SEQUENCE 1200 AA; 138973 MW; 616663FBA6B267D7 CRC64;

Query Match 89.6%; Score 86; DB 2; Length 1200;
 Best Local Similarity 90.0%; Pred. No. 0.013; 1; Indels 0; Gaps 0;
 Matches 18; Conservative 1; Mismatches 1;

OY 1 SNELTRAVEELHKLKEARE 20
 |||||:|||||
 Db 300 SNELTRAVDELHKLKEARE 319

RESULT 7
 Q7M739 PRELIMINARY; PRT; 2357 AA.
 ID Q7M739
 AC Q7M739
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Nuclear pore complex-associated intranuclear coiled-coil protein
 DE TPR.
 GN Name=Tpr;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22310828; PubMed=12424524;
 RA Kuznetsov N.V., Sandblad L., Hase M.E., Hunziker A., Hergt M.,
 RA Cordes V.C.;
 RT "The evolutionarily conserved single-copy gene for murine Tpr encodes
 RT one prevalent isoform in somatic cells and lacks paralogs in higher
 RT eukaryotes."
 RL Chromosome 11:236-255(2002).
 CC -!- MISCELLANEOUS: The sequence shown here is derived from an
 CC EMBL/GenBank/DBSJ third party annotation (TPA) entry.
 DR EMBL; BK00023; DAA00370.1; -.
 SQ SEQUENCE 2357 AA; 266924 MW; 3DA5A0FE2457453F CRC64;

Query Match 89.6%; Score 86; DB 2; Length 2357;
 Best Local Similarity 90.0%; Pred. No. 0.023; 1; Indels 0; Gaps 0;
 Matches 18; Conservative 1; Mismatches 1;

OY 1 SNELTRAVEELHKLKEARE 20
 |||||:|||||
 Db 300 SNELTRAVDELHKLKEARE 319

RESULT 8
 Q7P5L1 PRELIMINARY; PRT; 163 AA.
 ID Q7P5L1
 AC Q7P5L1
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE ATP synthase B chain, sodium ion specific (EC 3.6.3.15).
 GN Name=FN0984;
 OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.

OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium.
 OX NCBI_TaxID=209882;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 49256;
 RA Karpal V., Ivanova N., Anderson I., Reznik G., Bhattacharya A.,
 RA Gardner W.L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
 RA Haselkorn R., Overbeek R., Kyrpides N.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBSJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP + H₂O + H⁺ (In) = ADP + phosphate +
 CC H⁺ (Out).
 CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1) (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the ATPase B chain family.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAF01000062; EAA24012.1; -.
 DR GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.
 DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity; -; IEA.
 DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. -; IEA.
 DR GO; GO:0015787; F:hydrolase activity; IEA.
 DR GO; GO:0016820; F:hydrolase activity, acting on acid anhydrid. -; IEA.
 DR GO; GO:0015986; F:ATP synthetase coupled proton transport; IEA.
 DR InterPro; IPR002146; ATPsynt B/B' sub.
 DR InterPro; IPR005864; ATP synthase B.
 DR Pfam; PF00430; ATP-synt_B; 1.
 DR TIGRFAMs; TIGR01144; ATP synt_b; 1.
 KW CF(0): Hydrogen ion transport; Hydrolase; Transmembrane.
 SQ SEQUENCE 163 AA; 18894 MW; 92AF993E3718C92F CRC64;

Query Match 53.1%; Score 51; DB 2; Length 163;
 Best Local Similarity 66.7%; Pred. No. 76;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 6 RAVEELHKLKEARE 20
 :|||:|||||
 Db 85 KAEEEAHNLKEARE 99

RESULT 9
 Q73Z23 PRELIMINARY; PRT; 392 AA.
 ID Q73Z23
 AC Q73Z23
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocusNames=MAP1458;
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
 OC Mycobacterium avium complex (MAC).
 OX NCBI_TaxID=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=k10;
 RA Li L., Bannantine J., Zhang Q., Anonsin A., Alt D., Kapur V.;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBSJ databases.
 CC -!- SIMILARITY: Belongs to the acyl-CoA dehydrogenase family.
 DR EMBL; AE017232; AAS03775.1; -.
 DR InterPro; IPR006090; Acyl-CoA dh C.
 DR InterPro; IPR006091; Acyl-CoA dh M.
 DR InterPro; IPR006092; Acyl-CoA dh N.
 DR InterPro; IPR009075; AcylCoADH_C-like.
 DR InterPro; IPR009100; AcylCoA dehyd_NM.
 DR Pfam; PF00441; Acyl-CoA dh; 1.
 DR Pfam; PF02770; Acyl-CoA dh M; 1.
 DR Pfam; PF02771; Acyl-CoA dh N; 1.
 KW Complete proteome; FAD; Flavoprotein; Hypothetical protein;

KW Oxidoreductase.
SQ SEQUENCE 392 AA; 43695 MW; 1E7BA8EC5B045968 CRC64;

Query Match 53.1%; Score 51; DB 2; Length 392;
Best Local Similarity 52.6%; Pred. No. 1.7e+02;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKEAR 19
Db 250 AGDLRRRAVEELHKLKEAR 268

RESULT 10
AAS03775 PRELIMINARY; PRT; 392 AA.
ID AAS03775
AC AAS03775
DT 02-MAR-2004 (T-EMBLrel. 27, Created)
DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN MAP1458.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ki0;
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017232; AAS03775.1; -
KW Hypothetical protein.
SQ SEQUENCE 392 AA; 43695 MW; 1E7BA8EC5B045968 CRC64;

Query Match 53.1%; Score 51; DB 2; Length 392;
Best Local Similarity 52.6%; Pred. No. 1.7e+02;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 SNELTRAVEELHKLKEAR 19
Db 250 AGDLRRRAVEELHKLKEAR 268

RESULT 11
Q84YR9 PRELIMINARY; PRT; 418 AA.
ID Q84YR9
AC Q84YR9
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Putative Sec14 cytosolic factor (Phosphatidylinositol/phosphatidyl-choline transfer protein).
GN Name=OSJNB0027N13.111;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005641; BAC57373.1; -
DR Gramene; Q84YR9; -
DR InterPro; IPR001251; CRAL_TRIO_C.
DR InterPro; IPR008273; CRAL_TRIO_N.
DR InterPro; IPR011074; Sec14p-like_N.
DR Pfam; PF00650; CRAL_TRIO_1.
DR Pfam; PF03765; CRAL_TRIO_N; 1.
DR SMART; SMC0516; SEC14; 1.
DR PROSITE; PS0191; CRAL_TRIO; 1.
SQ SEQUENCE 418 AA; 47494 MW; EDF950C186EC6C98 CRC64;

Query Match 53.1%; Score 51; DB 2; Length 418;
Best Local Similarity 58.8%; Pred. No. 1.8e+02;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 NELTRAVEELHKLKEA 18
Db 389 NELANHIEQLRLKEA 405

RESULT 12
GCDA_ACIFE STANDARD; PRT; 587 AA.
ID GCDA_ACIFE
AC Q06700;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Glutaconyl-CoA decarboxylase alpha subunit (EC 4.1.1.70)
DE (Carboxyltransferase).
GN Name=gcda;
OS Acidaminococcus fermentans.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Acidaminococcaceae;
OC Acidaminococcus.
OX NCBI_TaxID=905;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 25085;
RX MEDLINE=93170302; PubMed=8382157;
RA Bendrat K., Buckel W.;
RT "Cloning, sequencing and expression of the gene encoding the carboxyltransferase subunit of the biotin-dependent Na⁺ pump glutaconyl-CoA decarboxylase from Acidaminococcus fermentans in Escherichia coli.";
RT Eur. J. Biochem. 211:697-702(1993).
RN [2]
RP SEQUENCE OF 548-587 FROM N.A.
RC STRAIN=ATCC 25085;
RX MEDLINE=93374040; PubMed=8365476;
RA Bendrat K., Mueller U., Klees A.G., Buckel W.;
RT "Identification of the gene encoding the activator of (R)-2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermentans by gene expression in Escherichia coli.";
RT FEBS Lett. 329:329-331(1993).
RN [3]
RP SEQUENCE OF 548-587 FROM N.A.
RC STRAIN=ATCC 25085;
RX MEDLINE=95331308; PubMed=7607244;
RA Mueller U., Buckel W.;
RT "Activation of (R)-2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermentans.";
RT Eur. J. Biochem. 230:698-704(1995).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=22737481; PubMed=12853465;
RA Wendt K.S., Schall I., Huber R., Buckel W., Jacob U.;
RT "Crystal structure of the carboxyltransferase subunit of the bacterial sodium ion pump glutaconyl-coenzyme A decarboxylase.";
RL EMBO J. 22:3493-3502(2003).
CC -!- FUNCTION: Decarboxylase subunit of the primary sodium pump glutaconyl-CoA decarboxylase (GCD).
CC -!- CATALYTIC ACTIVITY: 4-carboxybut-2-enoyl-CoA = but-2-enoyl-CoA + CO(2).
CC -!- PATHWAY: Anaerobic glutamate metabolism; fifth step.
CC -!- SUBUNIT: Heterotrimer consisting of two alpha, two beta, two gamma and two delta subunits.
CC -!- SIMILARITY: Contains 1 carboxyltransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

```

-----
CC EMBL; X69435; CAA49210.1; --
CC EMBL; X59645; CAA42195.1; --
DR PIR; S29787; S29787.
DR PDB; 1PIX; X-ray; A/B=1-587.
DR InterPro; IPR000022; Carboxyl trans.
DR Pfam; PF01039; Carboxyl trans; 1.
DR PROSITE; PS0989; COA_CT_CTER; 1.
DR PROSITE; PS0980; COA_CT_CTER; 1.
DR 3D-structure; Biotin; DeCarboxylase; Direct protein sequencing; Lyase;
KW Sodium transport. 568 Carboxyltransferase.
FT DOMAIN 35
SQ SEQUENCE 587 AA; 64346 MW; BF1C0D2F2DC8A6DA CRC64;

Query Match 53.1%; Score 51; DB 1; Length 587;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 ELTRAVERELHKLKEARE 20
DB 30 QLKKEEELHQLKEAQE 47

RESULT 13
Q9PE81 PRELIMINARY; PRT; 156 AA.
AC Q9PE81;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE ATP synthase B chain.
GN OrderedLocusNames=X1147;
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5c;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu P.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Collauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.B.S.,
RA Nhani A., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.B. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terezzi M.F., Trufi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zaco M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five

```

```

CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1) (By
CC similarity).
CC -!- SIMILARITY: Belongs to the ATPase B chain family.
DR EMBL; AE003950; AAF83957.1; --
DR PIR; C82716; C82716.
DR HSP; P00859; 1L2P.
DR GO; GO:0046469; C:proton-transporting two-sector ATPase complex; IEA.
DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity; IEA.
DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. ; IEA.
DR GO; GO:0046820; F:hydrolase activity, acting on acid anhydrid. ; IEA.
DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.
DR InterPro; IPR002146; ATPsynth_B/B'sub.
DR InterPro; IPR005864; ATP_synthF0_B.
DR Pfam; PF00430; ATP-synth_B; 1.
DR TIGRfams; TIGR01144; ATP_synth_b; 1.
KW CF(0); Complete proteome; Hydrogen ion transport; Transmembrane.
SQ SEQUENCE 156 AA; 17461 MW; 11169746B7ADE8AC CRC64;

Query Match 52.1%; Score 50; DB 2; Length 156;
Best Local Similarity 61.1%; Pred. No. 99;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 ELTRAVERELHKLKEARE 20
DB 53 ELAQAEIKKTLKNARE 70

RESULT 14
Q7UWL4 PRELIMINARY; PRT; 648 AA.
AC Q7UWL4;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=RS1950;
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294136; CAD72348.1; --
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 648 AA; 73073 MW; 5857A11AF4501263 CRC64;

Query Match 52.1%; Score 50; DB 2; Length 648;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 3 ELTRAVERELHKLKEARE 20
DB 413 ELRRADRLKILRLQRE 430

RESULT 15
RPOL METKA STANDARD; PRT; 112 AA.
AC Q8T281;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA-directed RNA polymerase subunit L (EC 2.7.7.6).
GN Name=rpol; OrderedLocusNames=MK0058;
OS Methanopyrus kandleri.

```

```

OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014; DOI=10.1073/pnas.032671499;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N)
CC -!- SIMILARITY: Belongs to the archaeobacteria RPOL / eukaryotic RPB11/
CC RFC19 RNA polymerase subunit family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE010306; AAM01275.1; -.
DR HAMAP; MF_00261; -.
DR InterPro; IPR008025; RNAP_RPB11-like.
DR InterPro; IPR008193; RNAP_RPB11_L.
DR Pfam; PF01193; RNA_pol_L; 1.
DR PROSITE; PS01154; RNA_POL_L_13KD; FALSE NEG.
KW Complete proteome; DNA-directed RNA polymerase; Transcription;
KW Transferase.
SQ SEQUENCE 112 AA; 13024 MW; C43CDB5BD988934 CRC64;

```

```

Query Match      51.0%; Score 49; DB 1; Length 112;
Best Local Similarity 58.8%; Pred.No. 99;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      4 LTRAVELHKLKEARE 20
Db      85 LRAIESELLEVESAKE 101

```

Search completed: November 10, 2004, 12:27:12
 Job time : 24.8827 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 9.8287 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-16

Perfect score: 178

Sequence: 1 TYNLLPPKRPKEVLTDFAKVLEKGVDSRS 33

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	173	97.2	989	3	US-09-110-517-4
2	53	29.8	88	4	US-09-583-110-4166
3	51.5	28.9	1022	1	US-08-271-364A-8
4	51.5	28.9	1022	2	US-08-222-715B-27
5	51	28.7	641	3	US-09-233-989-10
6	50	28.1	426	4	US-09-538-092-7
7	50	28.1	1377	4	US-09-711-164-467
8	49.5	27.8	595	4	US-09-640-419C-20
9	49.5	27.8	3077	6	5223423-2
10	49	27.5	145	4	US-09-270-767-35356
11	49	27.5	145	4	US-09-270-767-50573
12	49	27.5	155	4	US-09-134-000C-6609
13	49	27.5	468	4	US-09-328-352-6023
14	49	27.5	758	2	US-08-222-617A-6
15	49	27.5	3666	2	US-08-222-617A-12
16	49	27.5	3727	2	US-08-222-617A-27
17	49	27.5	3778	2	US-08-222-617A-2
18	48	27.0	187	4	US-09-248-796A-17518
19	48	27.0	388	3	US-08-861-774E-94
20	48	27.0	490	4	US-09-543-681A-7938
21	47.5	26.7	86	4	US-09-583-110-4921
22	47.5	26.7	215	2	US-08-659-251-8
23	47.5	26.7	215	3	US-09-256-490-8
24	47.5	26.7	215	4	US-09-489-039A-11817
25	47.5	26.7	215	5	PCT-US96-11445-8
26	47.5	26.7	304	4	US-09-107-532A-6460
27	47	26.4	233	4	US-09-252-991A-27758

28	47	26.4	381	4	US-09-328-352-7838	Sequence 7838, Ap
29	47	26.4	580	2	US-08-591-079-2	Sequence 2, Appli
30	47	26.4	608	4	US-08-637-670-36	Sequence 36, Appli
31	47	26.4	714	4	US-09-248-796A-18997	Sequence 18997, A
32	47	26.4	1007	2	US-08-551-459-4	Sequence 4, Appli
33	46.5	26.1	162	4	US-09-248-796A-32798	Sequence 22798, A
34	46.5	26.1	220	4	US-09-270-767-42243	Sequence 42243, A
35	46.5	26.1	220	4	US-09-270-767-57522	Sequence 57522, A
36	46.5	26.1	241	4	US-09-107-532A-4086	Sequence 4086, Ap
37	46.5	26.1	516	2	US-08-762-106-8	Sequence 8, Appli
38	46.5	26.1	516	3	US-08-745-404-2	Sequence 2, Appli
39	46.5	26.1	516	3	US-09-320-774-8	Sequence 8, Appli
40	46.5	26.1	527	2	US-08-762-106-9	Sequence 9, Appli
41	46.5	26.1	527	3	US-09-320-774-9	Sequence 9, Appli
42	46.5	26.1	552	3	US-08-745-404-3	Sequence 3, Appli
43	46.5	26.1	582	4	US-09-543-681A-4556	Sequence 4556, Ap
44	46.5	26.1	606	4	US-09-252-991A-22263	Sequence 22263, A
45	46.5	26.1	1683	3	US-08-755-587-183	Sequence 183, App

ALIGNMENTS

RESULT 1

US-09-110-517-4
; Sequence 4, Application US/09110517A
; Patent No. 6248520
; GENERAL INFORMATION:
; APPLICANT: Roeder, Robert G
; APPLICANT: Fondell, Joseph D
; APPLICANT: Yuan, Chao X
; APPLICANT: Ito, Mitsuhiro
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NUCLEAR HORMONE
; TITLE OF INVENTION: RECEPTOR COACTIVATORS AND USES THEREOF
; FILE REFERENCE: 600-1-224
; CURRENT APPLICATION NUMBER: US/09/110,517A
; CURRENT FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 989
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-110-517-4

Query Match 97.2%; Score 173; DB 3; Length 989;
Best Local Similarity 100.0%; Pred. No. 1.4e-16;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 YWNLPPKRPKEVLTDFAKVLEKGVDSRS 33

DB 697 YWNLPPKRPKEVLTDFAKVLEKGVDSRS 728

RESULT 2

US-09-583-110-4166
; Sequence 4166, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: P400-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4166

LENGTH: 88
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-583-110-4166

Query Match 29.8%; Score 53; DB 4; Length 88;
Best Local Similarity 45.0%; Pred. No. 2.4;
Matches 9; Conservative 6; Mismatches 5; Indels 5; Gaps 0;

QY 9 KRIKEVLTDFIAKVLKGM 28
DB 17 KKEISETLTDVVASLNDKGY 36

RESULT 3

US-08-271-364A-8
Sequence 8, Application US/08271364A
Patent No. 5756334

GENERAL INFORMATION:

APPLICANT: PERLER, FRANCINE B.
APPLICANT: SOUTHWORTH MAURICE W.
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA POLYMERASE
TITLE OF INVENTION: FROM ARCHAEBACTERIA
NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSER: NEW ENGLAND BIOLABS, INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 01915

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/271,364A
FILING DATE: 06-JUL-1994

CLASSIFICATION:

435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/811,421
FILING DATE: 18-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/686,340
FILING DATE: 17-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057
FILING DATE: 11-DEC-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/513,994
FILING DATE: 26-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, GREGORY D.

REGISTRATION NUMBER: 30901

REFERENCE/DOCKET NUMBER: NEB-101

TELEPHONE: (508) 927-5054

TELEFAX: (508) 927-1705

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 1022 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-271-364A-8

Query Match 28.9%; Score 51.5; DB 1; Length 1022;
Best Local Similarity 44.0%; Pred. No. 66;
Matches 11; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

QY 2 YWNLLPPKPIKEVLTDFIAKVLK 26

DB 942 YSNLIP-----KEILRDVFGKGFQK 961

RESULT 4

US-08-222-715B-27
Sequence 27, Application US/08222715B
Patent No. 5834285

GENERAL INFORMATION:

APPLICANT: Comb, Donald G.
APPLICANT: Perler, Francine
APPLICANT: Kucera, Rebecca
APPLICANT: Jack, William E.
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA
TITLE OF INVENTION: POLYMERASE FROM ARCHAEBACTERIA
NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSER: GREGORY D. WILLIAMS; NEW ENGLAND BIOLABS,
ADDRESSER: INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 01915

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,715B
FILING DATE: 04-APR-1994

CLASSIFICATION:

536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/167,238
FILING DATE: 15-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/686,340
FILING DATE: 17-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057
FILING DATE: 11-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/513,994
FILING DATE: 26-APR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Williams, Gregory D.

REGISTRATION NUMBER: 30901

REFERENCE/DOCKET NUMBER: NEB-054C3FC2

TELEPHONE: (508) 927-5054

TELEFAX: (508) 927-1705

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 1022 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-222-715B-27

Query Match 28.9%; Score 51.5; DB 2; Length 1022;
Best Local Similarity 44.0%; Pred. No. 66;
Matches 11; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

QY 2 YWNLLPPKPIKEVLTDFIAKVLK 26

DB 942 YSNLIP-----KEILRDVFGKGFQK 961

RESULT 5

US-09-233-989-10


```

; Sequence 10, Application US/09233989
; Patent No. 6248527
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Meyer, Joanne
; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on
; FILE REFERENCE: 5800-14, 035800/174130
; CURRENT APPLICATION NUMBER: US/09/233,989
; CURRENT FILING DATE: 1999-01-19
; EARLIER APPLICATION NUMBER: 60/105,102
; EARLIER FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: carboxypeptidase homolog -- CPZ
US-09-233-989-10

Query Match      28.7%; Score 51; DB 3; Length 641;
Best Local Similarity 38.7%; Pred. No. 46;
Matches 12; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

Qy  2 YWLLPPKPIKEVLTDFIAKLEKGWVDSR 32
Db  529 YWLLPPGIHIVIAQAAGYAKVIKKVIIPAR 559

RESULT 6
US-09-538-092-7
; Sequence 7, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormater Version 0.9
; SEQ ID NO 7
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YAR003W
US-09-538-092-7

Query Match      28.1%; Score 50; DB 4; Length 426;
Best Local Similarity 30.0%; Pred. No. 40;
Matches 9; Conservative 9; Mismatches 10; Indels 2; Gaps 1;

Qy  3 WNLPPKPIKEVLTDFIAKLEKGWVDSR 32
Db  99 WDLSPKSLKEIRFD--SPIWGQWLDAK 126

RESULT 7
US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn

```

```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA 008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 1377
; TYPE: PRT
; ORGANISM: Escherichia coli
; OTHER INFORMATION:
US-09-711-164-467

Query Match      28.1%; Score 50; DB 4; Length 1377;
Best Local Similarity 42.3%; Pred. No. 1.5e+02;
Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy  8 PKRPIKEVLTDFIAKLEKGWVDSRS 33
Db  1261 PKSPAQQQADDDNAKALTQWWDITAS 1286

RESULT 8
US-09-640-419C-20
; Sequence 20, Application US/09640419C
; Patent No. 6630615
; GENERAL INFORMATION:
; APPLICANT: Bidney, Dennis L
; APPLICANT: Craata, Oswald R
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guohua
; TITLE OF INVENTION: DEFENSE-RELATED SIGNALING GENES AND METHODS OF USE
; FILE REFERENCE: 35718/199009 (5718-92)
; CURRENT APPLICATION NUMBER: US/09/640,419C
; CURRENT FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/149,656
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/206,405
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; OTHER INFORMATION:
US-09-640-419C-20

Query Match      27.8%; Score 49.5; DB 4; Length 595;
Best Local Similarity 52.2%; Pred. No. 69;
Matches 12; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

Qy  7 PPKRPI---KEVLTDFIAKLEK 26
Db  99 PPSRPSVDPKHYLSNDNFAPVLDE 121

RESULT 9
5223423-2
; Patent No. 5223423
; APPLICANT: FRANCHINI, GENOVEFFA;WONG-STAAAL, FLOSSIE;
; GALLO, ROBERT
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/331,212
; FILING DATE: 03-31-1989
; SEQ ID NO:2:
; LENGTH: 3077
5223423-2

```

```

Query Match      27.8%; Score 49.5; DB 6; Length 3077;
Best Local Similarity 34.4%; Pred. No. 4.5e+02;
Matches 11; Conservative 1; Mismatches 3; Indels 17; Gaps 1;

QY 2 YWNLPPRRPIKEVLTDFIAKVLKGVWDSRS 33
    |||||
Db 1585 YWNLTP-----EKGWLSYS 1599

RESULT 10
US-09-270-767-35356
; Sequence 35356, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35356
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-35356

Query Match      27.5%; Score 49; DB 4; Length 145;
Best Local Similarity 45.5%; Pred. No. 16;
Matches 10; Conservative 5; Mismatches 7; Indels 7; Gaps 0;

QY 4 NLLPPKRPKEVLTDFIAKVL 25
    |||||
Db 112 NSLXPNMKIEKVTKVPAKCLQ 133

RESULT 11
US-09-270-767-50573
; Sequence 50573, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50573
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-50573

Query Match      27.5%; Score 49; DB 4; Length 145;
Best Local Similarity 45.5%; Pred. No. 16;
Matches 10; Conservative 5; Mismatches 7; Indels 7; Gaps 0;

QY 4 NLLPPKRPKEVLTDFIAKVL 25
    |||||
Db 112 NSLXPNMKIEKVTKVPAKCLQ 133

RESULT 12
US-09-134-000C-6609
; Sequence 6609, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6609
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6609

Query Match      27.5%; Score 49; DB 4; Length 155;
Best Local Similarity 40.9%; Pred. No. 18;
Matches 9; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 8 PKRPIKEVLTDFIAKVLKGVW 29
    |||||
Db 100 PEKPFVDVIITDIFSEVSVKSTV 121

RESULT 13
US-09-328-352-6023
; Sequence 6023, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6023
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6023

Query Match      27.5%; Score 49; DB 4; Length 468;
Best Local Similarity 40.9%; Pred. No. 62;
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 8 PKRPIKEVLTDFIAKVLKGVW 29
    |||||
Db 289 PYEPIIDKAMEIFHCMEQGVW 310

RESULT 14
US-08-222-617A-6
; Sequence 6, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; NUMBER OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive

```

RESULT 15
 US-08-222-617A-12
 ; Sequence 12, Application US/08222617A
 ; Patent No. 5882879
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Veenstra, Annemarie E.
 ; APPLICANT: Martin, Juan P.
 ; APPLICANT: Garcia, Bruno D.
 ; APPLICANT: Gutierrez, Santiago
 ; APPLICANT: Barredo, Jose L.
 ; APPLICANT: Von Doehren, Hans
 ; APPLICANT: Palissa, Harriet
 ; APPLICANT: Van Liempt, Henk
 ; APPLICANT: Montenegro, Eduardo P.
 ;
 ; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
 ; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
 ; TITLE OF INVENTION: Quantities of ACV Synthetase
 ;
 ; NUMBER OF SEQUENCES: 27
 ;
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
 ; STREET: 300 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: USA
 ;
 ; ZIP: 60606
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ;
 ; CURRENT APPLICATION DATA:

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 15:53:52 ; Search time 31.8287 Seconds
(without alignments)
366.225 Million cell updates/sec

Title: US-10-092-750-16

Perfect score: 178
Sequence: 1 TYWNLLPPKRPKEVLTIDIFAKVLEKGMWDSRS 33

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	178	100.0	33	14	US-10-092-750-16
2	173	97.2	989	15	US-10-435-696-40
3	65.5	36.8	168	16	US-10-767-701-54304
4	62	34.8	102	16	US-10-437-963-112003
5	57	32.0	170	15	US-10-424-599-171900
6	56	31.5	341	15	US-10-424-599-271602
7	55.5	31.2	277	14	US-10-369-493-14954
8	55.5	31.2	282	14	US-10-369-493-14462
9	55.5	31.2	495	14	US-10-369-493-11629
10	55.5	31.2	985	16	US-10-437-963-162348
11	55	30.9	109	17	US-10-425-115-267928
12	55	30.9	329	16	US-10-437-963-202829
13	53.5	30.1	150	16	US-10-437-963-121135

14	53	29.8	88	9	US-09-815-242-13194	Sequence 1194, A
15	53	29.8	88	15	US-10-282-122A-73718	Sequence 73718, A
16	53	29.8	95	16	US-10-437-963-173316	Sequence 173316, A
17	52.5	29.5	303	14	US-10-032-214-205	Sequence 205, App
18	52.5	29.5	303	14	US-10-032-214-285	Sequence 285, App
19	52.5	29.5	1199	16	US-10-437-963-121132	Sequence 121132, A
20	52	29.2	217	15	US-10-424-599-246489	Sequence 246489, A
21	52	29.2	722	15	US-10-282-122A-67796	Sequence 67796, A
22	51.5	28.9	73	16	US-10-437-963-146540	Sequence 146540, A
23	51.5	28.9	856	16	US-10-437-963-148083	Sequence 148083, A
24	51	28.7	85	17	US-10-425-115-257431	Sequence 257431, A
25	51	28.7	86	16	US-10-437-963-163078	Sequence 163078, A
26	51	28.7	88	16	US-10-437-963-103242	Sequence 103242, A
27	51	28.7	155	15	US-10-282-122A-77684	Sequence 77684, A
28	51	28.7	641	9	US-09-827-040-10	Sequence 10, Appl
29	51	28.7	641	17	US-10-757-262-104	Sequence 104, Appl
30	51	28.7	662	15	US-10-425-114-43276	Sequence 43276, A
31	51	28.7	888	15	US-10-425-114-45466	Sequence 45466, A
32	51	28.7	967	17	US-10-425-115-194555	Sequence 194555, A
33	51	28.7	970	15	US-10-425-114-52042	Sequence 52042, A
34	51	28.7	970	15	US-10-425-114-58607	Sequence 58607, A
35	51	28.7	1472	17	US-10-425-115-316823	Sequence 316823, A
36	50.5	28.4	145	15	US-10-424-599-252168	Sequence 252168, A
37	50.5	28.4	429	16	US-10-437-963-190212	Sequence 190212, A
38	50.5	28.4	537	15	US-10-389-566-838	Sequence 838, App
39	50.5	28.4	598	16	US-10-437-963-154209	Sequence 154209, A
40	50.5	28.4	614	15	US-10-389-566-839	Sequence 839, App
41	50.5	28.4	1286	16	US-10-437-963-121139	Sequence 121139, A
42	50	28.1	61	16	US-10-767-701-48231	Sequence 48231, A
43	50	28.1	518	16	US-10-686-947-298	Sequence 298, App
44	50	28.1	972	16	US-10-437-963-154999	Sequence 154999, A
45	50	28.1	1377	9	US-09-815-242-10384	Sequence 10384, A

ALIGNMENTS

RESULT 1
US-10-092-750-16
; Sequence 16, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCU-XI
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; TYPE: PRT
; LENGTH: 33
; ORGANISM: Homo sapiens
US-10-092-750-16

Query Match	100.0%	Score 178:	DB 14:	Length 33;
Best Local Similarity	100.0%	Pred. No. 3.4e-17;		
Matches 33;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	TYWNLLPPKRPKEVLTIDIFAKVLEKGMWDSRS 33		
Db	1	TYWNLLPPKRPKEVLTIDIFAKVLEKGMWDSRS 33		
RESULT 2				
US-10-435-696-40				
; Sequence 40, Application US/10435696				
; Publication No. US20040018525A1				
; GENERAL INFORMATION:				

```

; APPLICANT: Wirtz, Ralph
; APPLICANT: Munnes, Marc
; APPLICANT: Kallabis, Harald
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE PREDICTION, DIAGNOSIS, PROGNOSIS
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF MALIGNANT NEOPLASIA
; FILE REFERENCE: LEA 36 108
; CURRENT APPLICATION NUMBER: US/10/435.696
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: EP03003112.4
; PRIOR FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: EP02010291.9
; PRIOR FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 314
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 40
; LENGTH: 989
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-435-696-40

```

```

Query Match 97.2%; Score 173; DB 15; Length 989;
Best Local Similarity 100.0%; Pred. No. 7e-15;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 YWNLPPKRPKEVLTDFIAKVGWVDSRS 33
DB 697 YWNLPPKRPKEVLTDFIAKVGWVDSRS 728

```

```

RESULT 3
US-10-767-701-54304
; Sequence 54304, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53335)B
; CURRENT APPLICATION NUMBER: US/10/767.701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 54304
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 4513462.pap
US-10-767-701-54304

```

```

Query Match 36.8%; Score 65.5; DB 16; Length 168;
Best Local Similarity 40.0%; Pred. No. 0.67;
Matches 14; Conservative 6; Mismatches 10; Indels 5; Gaps 1;

```

```

QY 2 YWNLPPKRPKEVLTDFIAKV-----LEKGWVDS 31
DB 98 YWGLHPPKPLEEVIDDGLVAVQOALELKKGVDT 132

```

```

RESULT 4
US-10-437-963-112003
; Sequence 112003, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping

```

```

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437.963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 112003
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(102)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_1592C.1.pap
US-10-437-963-112003

```

```

Query Match 34.8%; Score 62; DB 16; Length 102;
Best Local Similarity 40.0%; Pred. No. 1.2;
Matches 10; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 2 YWNLPPKRPKEVLTDFIAKVGLEK 26
DB 27 YWRLPPRLVLEELPGDLSEKIVEE 51

```

```

RESULT 5
US-10-424-599-171900
; Sequence 171900, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 171900
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_12623C.1.pap
US-10-424-599-171900

```

```

Query Match 32.0%; Score 57; DB 15; Length 170;
Best Local Similarity 45.5%; Pred. No. 10;
Matches 15; Conservative 5; Mismatches 9; Indels 4; Gaps 2;

```

```

QY 1 TYWNL-----PPKRPKEVLTDFIAKVGKWD 30
DB 81 THYNQLATPAARPLTVFSDIFERV-KGWKD 112

```

```

RESULT 6
US-10-424-599-271602
; Sequence 271602, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424.599
; CURRENT FILING DATE: 2003-04-28

```

; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 271602
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87276C.1.pap
US-10-424-599-271602

Query Match 31.5%; Score 56; DB 15; Length 341;
Best Local Similarity 34.4%; Pred. No. 30;
Matches 11; Conservative 6; Mismatches 11; Indels 4; Gaps 1;

QY 2 YNLLPKPIKEVLTDIFAKVLEKGVDSRS 33
||| : : : : :
Db 28 YNKSVLPTTFMKAITDI----LYSDWVEKS 55

RESULT 7
US-10-369-493-14954
; Sequence 14954, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14954
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14954

Query Match 31.2%; Score 55.5; DB 14; Length 277;
Best Local Similarity 42.9%; Pred. No. 28;
Matches 15; Conservative 2; Mismatches 9; Indels 9; Gaps 2;

QY 5 LLPPKR-----PIKEVLTDIFAKVLE--KGWVD 30
||| : : : : :
Db 113 LLPTRTDLRKDPAGKPLKPVFAKFEYSDGWVD 147

RESULT 8
US-10-369-493-14462
; Sequence 14462, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14462
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens

US-10-369-493-14462

Query Match 31.2%; Score 55.5; DB 14; Length 282;
Best Local Similarity 42.9%; Pred. No. 28;
Matches 15; Conservative 2; Mismatches 9; Indels 9; Gaps 2;

QY 5 LLPPKR-----PIKEVLTDIFAKVLE--KGWVD 30
||| : : : : :
Db 118 LLPTRTDLRKDPAGKPLKPVFAKFEYSDGWVD 152

RESULT 9
US-10-369-493-11629
; Sequence 11629, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11629
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11629

Query Match 31.2%; Score 55.5; DB 14; Length 495;
Best Local Similarity 42.9%; Pred. No. 53;
Matches 15; Conservative 2; Mismatches 9; Indels 9; Gaps 2;

QY 5 LLPPKR-----PIKEVLTDIFAKVLE--KGWVD 30
||| : : : : :
Db 117 LLPTRTDLRKDPAGKPLKPVFAKFEYSDGWVD 151

RESULT 10
US-10-437-963-162348
; Sequence 162348, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162348
; LENGTH: 985
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_61448C.1.pap
US-10-437-963-162348

Query Match 31.2%; Score 55.5; DB 16; Length 985;
Best Local Similarity 37.1%; Pred. No. 1.1e+02;

Matches 13; Conservative 6; Mismatches 11; Indels 5; Gaps 1;

QY 2 YNLLPPRKPIKEVLTDF----AKVLKGVWDS 31
||| |:::||: ||: ||:
Db 411 YWGLHAPKPLEEIVDDGLVTLOQLFLKVAGVD 445

```

RESULT 11
US-10-425-115-267928
; Sequence 267928, Application US/10425115
; Publication NO. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid
; MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 267928
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_175952C.1.pep
US-10-425-115-267928

```

Qy 6 LPPKRPKEVLTDIFAKVLEKG 27
 :|||||: | : | : |||:
Db 52 VPKRVPPIALENSWASLLERG 73

```

RESULT 12
US-10-437-963-202829
; Sequence 202829, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221) B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 202829
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MFT4530_98071C.1 pep
US-10-437-963-202829

```

QY 3 W N L L P P K R P I K E V L T D I F A K V L E X 26
 | | | | : : : : :
DB 22 W R L P P P R R L V E E L P G D L S E K I V E E 45

RESULT 13
US-10-437-963-121135
; Sequence 121135, Application US/10437963
; Publication No. US20040123343A1

```

QY      2  YWLLPPKPIKEVLTDIFA---KVLEKG-----WVDSRS 33
      ||:| | | | | | | | | | | | | | | | | | | | |
Db     55  YWSDL--KAPIKRSVDMFVAYTFSILGDGOSKLEWIDRWIDGRS 97

```

```

RESULT 14
US-09-815-242-13194
// Sequence 13194, Application US/09815242
// Patent No. US20020061569A1
// GENERAL INFORMATION:
// APPLICANT: Haselbeck, Robert
// APPLICANT: Chlsen, Kari L.
// APPLICANT: Zyskind, Judith W.
// APPLICANT: Wall, Daniel
// APPLICANT: Trawick, John D.
// APPLICANT: Carr, Grant J.
// APPLICANT: Yamamoto, Robert T.
// APPLICANT: Xu, H. Howard
// TITLE OF INVENTION: Identification of Essential Genes in
// TITLE OF INVENTION: Prokaryotes
// FILE REFERENCE: ELITRA-011A
// CURRENT APPLICATION NUMBER: US/09/815,242
// CURRENT FILING DATE: 2001-03-21
// PRIOR APPLICATION NUMBER: 60/191,078
// PRIOR FILING DATE: 2000-03-21
// PRIOR APPLICATION NUMBER: 60/206,848
// PRIOR FILING DATE: 2000-05-23
// PRIOR APPLICATION NUMBER: 60/207,727
// PRIOR FILING DATE: 2000-05-26
// PRIOR APPLICATION NUMBER: 60/242,578
// PRIOR FILING DATE: 2000-10-23
// PRIOR APPLICATION NUMBER: 60/253,625
// PRIOR FILING DATE: 2000-11-27
// PRIOR APPLICATION NUMBER: 60/257,931
// PRIOR FILING DATE: 2000-12-22
// PRIOR APPLICATION NUMBER: 60/269,308
// PRIOR FILING DATE: 2001-02-16
// NUMBER OF SEQ ID NOS: 14110
// SOFTWARE: FastSeq for Windows Version 4.0
// SEQ ID NO 13194

```


Search completed: November 10, 2004, 16:36:00
Job time : 32.8287 secs

```

; LENGTH: 88
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13194

```

Query Match 29.8%; Score 53; DB 9; Length 88;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy	9	KRPIKEVLTDIFAKVLEKGW	28
		: : :	
Db	17	KKEISETLTDVYASLNDKGY	36

RESULT 15
US-10-282-122A-73718
; Sequence 73718, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```

/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 73718
/ LENGTH: 86
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
/ US-10-282-122A-73718

```

```
Query Match      29.8%; Score 53; DB 15; Length 88;
Best Local-Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
```

Qy 9 KRPIKEVLTDIFAKVLEKGW 28
| : | | | | : : | : | :
Db 17 KKEISETLTDVYASLNDKGY 36

This page blank (uspio)

Result No.	Score	Query Match	Length	DB	ID	Description
1	57	32.0	314	2	S38165	hypothetical prote
2	56	31.5	262	2	D69438	hypothetical prote
3	55.5	31.2	515	2	AH2996	glycerol-3-phospha
4	55.5	30.9	531	2	A98287	glpD gene homolog
5	55	30.9	529	2	S58155	RNA binding protei
6	53	29.8	88	2	D95022	conserved hypoteth
7	53	29.8	88	2	G97893	hypothetical prote
8	52	29.2	164	2	B37842	hypothetical prote
9	52	29.2	192	2	AF1987	hypothetical prote
10	52	29.2	457	2	S44269	platelet-derived g
11	52	29.2	1461	2	B70588	Probable polyketid
12	51.5	28.9	215	1	AS1JGG	vif protein - huma
13	51.5	28.9	215	2	S12154	vif protein - huma
14	51.5	28.9	1702	2	S43459	DNA-directed DNA p
15	51	28.7	155	2	A82390	conserved hypoteth
16	51	28.7	960	2	S18240	phosphoenolpyruvat
17	51	28.7	964	1	QVNT	phosphoenolpyruvat
18	51	28.7	966	2	S28614	phosphoenolpyruvat
19	51	28.7	967	2	JH0667	phosphoenolpyruvat
20	51	28.7	1278	1	E93615	enterobactin synth
21	50	28.1	181	2	I69096	cdtC protein - Esc
22	50	28.1	426	2	S40901	FUN16 protein - ye
23	50	28.1	771	2	H69305	signal-transducing
24	50	28.1	963	2	S49344	phosphoenolpyruvat
25	50	28.1	966	2	S13318	phosphoenolpyruvat
26	50	28.1	966	2	S37072	phosphoenolpyruvat
27	50	28.1	967	2	S25082	phosphoenolpyruvat
28	50	28.1	1377	2	E86034	rheA protein in rh
29	50	28.1	1377	2	C65159	rheA protein in presu

A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: D69438
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-262 <KLE>
 A:Cross-references: UNIPROT:O28763; GB:AE000998; GB:AE000782; NID:g2689321; PIDN:AA88974

Query Match 31.5%; Score 56; DB 2; Length 262;
 Best Local Similarity 44.0%; Pred. No. 5.4;
 Matches 11; Conservative 5; Mismatches 7; Indels 2; Gaps 1;

QY 1 TYNNLPPKPIKEVLTDIFAKVLE 25
 ||||| : : : : :
 190 TYNFFPPKPIQ-MTDAGTVE 212

Db 121 LLPTRTLDLRKDPAGKPLKPVFAKAFESDGVWD 155

RESULT 4
 A98287
 glpD gene homolog [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C:Accession: A98287
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, D.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2338, 2001
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: A98287
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-531 <KUR>
 A:Cross-references: UNIPROT:Q8U927; GB:AE007870; PIDN:AAK89819.1; PID:g15159751; GSPDB:G
 C:Genetics:
 A:Map position: linear chromosome
 A:Gene: AGR_L_2503
 C:Superfamily: glycerol-3-phosphate dehydrogenase (aerobic)

Query Match 31.2%; Score 55.5; DB 2; Length 531;
 Best Local Similarity 42.9%; Pred. No. 14;
 Matches 15; Conservative 2; Mismatches 9; Indels 9; Gaps 2;

QY 5 LLPKPR-----PIKEVLTDIFAKVLE--KGWVD 30
 ||||| : : : : :
 121 LLPTRTLDLRKDPAGKPLKPVFAKAFESDGVWD 155

Db 121 LLPTRTLDLRKDPAGKPLKPVFAKAFESDGVWD 155

RESULT 3
 AH2996
 glycerol-3-phosphate dehydrogenase glpD [imported] - Agrobacterium tumefaciens (strain C58)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
 C:Accession: AH2996
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, K.; Karp, P.; Romero, P.; Zhang, S.; Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AH2996
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-515 <KUR>
 A:Cross-references: UNIPROT:Q8U927; GB:AE008689; PIDN:AAL44390.1; PID:g17741987; GSPDB:G
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Map position: linear chromosome
 A:Gene: glpD
 C:Superfamily: glycerol-3-phosphate dehydrogenase (aerobic)

Query Match 31.2%; Score 55.5; DB 2; Length 515;
 Best Local Similarity 42.9%; Pred. No. 13;
 Matches 15; Conservative 2; Mismatches 9; Indels 9; Gaps 2;

QY 5 LLPKPR-----PIKEVLTDIFAKVLE--KGWVD 30
 ||||| : : : : :
 121 LLPTRTLDLRKDPAGKPLKPVFAKAFESDGVWD 155

Db 121 LLPTRTLDLRKDPAGKPLKPVFAKAFESDGVWD 155

Matches 15; Conservative 2; Mismatches 9; Indels 9; Gaps 2;

QY 5 LLPKPR-----PIKEVLTDIFAKVLE--KGWVD 30
 ||||| : : : : :
 137 LLPTRTLDLRKDPAGKPLKPVFAKAFESDGVWD 171

Db 137 LLPTRTLDLRKDPAGKPLKPVFAKAFESDGVWD 171

RESULT 5
 S58155
 RNA binding protein Nrd1 [validated] - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
 A:Accession: T38559; T50458; S58155
 R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V. submitted to the EMBL Data Library, July 1995
 A:Reference number: Z21799
 A:Accession: T38559
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-529 <GE2>
 A:Cross-references: UNIPROT:Q09702; EMBL:Z50142; NID:g1052783; PIDN:CAA90498.1; PID:g1052783
 A:Experimental source: strain 972h(-); cosmid c2F7
 R:Tsukahara, K.; Yamamoto, H.; Okayama, H.; Mol. Cell. Biol. 18, 4488-4498, 1998
 A>Title: An RNA binding protein negatively controlling differentiation in fission yeast.
 A:Reference number: Z25070; MUID:98336254; PMID:9671458
 A:Accession: T50458
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-529 <TSU>
 A:Cross-references: EMBL:AF079876; NID:g3396074; PIDN:AAC28857.1; PID:g3396075
 A:Experimental source: strain L972
 C:Genetics:
 A:Gene: SPAC2F7.11; nrd1+
 A:Map position: 1
 C:Function:
 A>Description: functions as negative regulator of differentiation; blocks the onset of sexual differentiation; blocks the onset of sporulation
 A>Note: preferentially binds poly(U)
 A:Note: Domain: ribonucleoprotein repeat homology <RM3>
 F:207-265/Domain: ribonucleoprotein repeat homology <RM3>

Query Match 30.9%; Score 55; DB 2; Length 529;
 Best Local Similarity 35.7%; Pred. No. 16;
 Matches 10; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 2 YNNLPPKPIKEVLTDIFAKVLEKGV 29
 ||||| : : : : :
 118 YVGNLPPNTPIDEILSCVRTGPIESAWI 145

Db 118 YVGNLPPNTPIDEILSCVRTGPIESAWI 145

RESULT 6
 D95022
 conserved hypothetical protein SP0192 [imported] - Streptococcus pneumoniae (strain TIGR)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
 C:Accession: D95022
 R:Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide, O.; J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle, I.; neon, T.; Hickey, E.K.; Holt, I.E. Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: D95022
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-88 <KUR>
 A:Cross-references: UNIPROT:Q97SX1; GB:AE005672; PIDN:AAK74373.1; PID:g14971660; GSPDB:GN
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP0192

Query Match 29.8%; Score 53; DB 2; Length 88;

```

Best Local Similarity 45.0%; Pred. No. 4.1;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 9 KRPIKEVLTDIFAKVLEKGW 28
| : | | | | : : | | :
Db 17 KKEISETLTDVYASLNDKGY 36

RESULT 7
G97893
hypothetical protein spr0175 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: G97893
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; BURGERT, S.; DeHoff, B.S.; E
Y, P.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G97893
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-88 <KUR>
A:Cross-references: UNIPROT:Q8CZ89; GB:AE007317; PIDN:AAK98979.1; PID:G15457718; GSPDB:G
C:Genetics:
A:Gene: spr0175

Query Match 29.8%; Score 53; DB 2; Length 88;
Best Local Similarity 45.0%; Pred. No. 4.1;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 9 KRPIKEVLTDIFAKVLEKGW 28
| : | | | | : : | | :
Db 17 KKEISETLTDVYASLNDKGY 36

RESULT 8
B37842
hypothetical protein 2 (xisa 3' region) - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
C:Date: 21-Jun-1991 #sequence_revision 21-Jun-1991 #text_change 09-Jul-2004
C:Accession: B37842
R:Hammers, P.J.; McLaughlin, S.; Papin, S.; Trujillo-Provencio, C.; Ryncarz II, A.J.
J. Bacteriol. 172, 6981-6990, 1990
A:Title: Developmental rearrangement of cyanobacterial nif genes: nucleotide sequence, c
A:Reference number: A37842; MUID:91072249; PMID:2123860
A:Accession: B37842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-164 <LAM>
A:Cross-references: UNIPROT:P29979; GB:U38537; GB:M38044; NID:G1053082; PIDN:AAC82966.1;
C:Superfamily: Anabaena hypothetical protein 2 (xisa 3' region)

Query Match 29.2%; Score 52; DB 2; Length 164;
Best Local Similarity 37.9%; Pred. No. 11;
Matches 11; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 3 WNLLPPKRPKEVLTDIFAKVLEKGWDS 31
| | | | | : : | | :
Db 55 WTLFPLKSTLRKVVDDGRDVIKGEIIS 83

RESULT 9
AF1987
hypothetical protein alr1449 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AF1987
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

```

```

DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF1987
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <KUR>
A:Cross-references: UNIPROT:P29979; GB:BA000019; PIDN:BAW73406.1; PID:G17130796; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
C:Superfamily: Anabaena hypothetical protein 2 (xisa 3' region)

Query Match 29.2%; Score 52; DB 2; Length 192;
Best Local Similarity 37.9%; Pred. No. 13;
Matches 11; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 3 WNLLPPKRPKEVLTDIFAKVLEKGWDS 31
| | | | | : : | | :
Db 83 WTLFPLKSTLRKVVDDGRDVIKGEIIS 111

RESULT 10
S44269
platelet-derived growth factor receptor alpha - zebra fish (fragment)
N:Contains: protein-tyrosine kinase (SC 2.7.1.112)
C:Species: Brachydanio rerio (zebra fish)
C:Date: 13-Jan-1995 #sequence_revision 07-Feb-1997 #text_change 09-Jul-2004
A:Accession: S44269
R:Yu, W.; Richardson, W.D.
submitted to the EMBL Data Library, April 1994
A:Description: Zebrafish Platelet-derived Growth Factor Receptor (Alpha subunit), partial
A:Reference number: S44269
A:Accession: S44269
A:Molecule type: mRNA
A:Residues: 1-457 <YUM>
A:Cross-references: UNIPROT:Q90269; EMBL:Z32814; NID:G479105; PIDN:CAA83675.1; PID:G47910
C:Superfamily: macrophage colony-stimulating factor 1 receptor; immunoglobulin homology;
C:Keywords: ATP; autophosphorylation; dimer; glycoprotein; growth factor receptor; phospho
P.1-327/Domain: protein kinase homology (fragment) <KIN>

Query Match 29.2%; Score 52; DB 2; Length 457;
Best Local Similarity 42.3%; Pred. No. 35;
Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 3 WNLLPPKRPKEVLTDIFAKVLEKGW 28
| | | | | : : | | :
Db 303 WNSEPEKRPFSHLSDTVASLLPSGF 328

RESULT 11
B70588
probable polyketide synthetase mbtF [similarity] - Mycobacterium tuberculosis (strain H37
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: B70588
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome s
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70588
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1461 <COL>
A:Cross-references: UNIPROT:O05819; GB:Z95208; GB:AL123456; NID:G3261747; PIDN:CA508474.1
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: mbtF
C:Superfamily: Mycobacterium tuberculosis mbtF protein; acetate-CoA ligase homology; acyl
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein

```


A;Gene: VCA1012
 A;Map position: 2
 C;Superfamily: spore maturation protein, SpMB type

Query Match	28.7%	Score 51;	DB 2;	Length 155;
Best Local Similarity	52.9%	Pred. No. 15;		
Matches	9;	Conservative	4;	Mismatches 4;
				Indels 0;
				Gaps 0;

QY 13 KEVLTDFIAKVKLGWV 29
 | : ||| : |||
 Db 8 KPMVTDIFVEGAKKGWV 24

Search completed: November 10, 2004, 12:29:20
 Job time : 7.6713 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01 ; Search time 36.1065 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-16

Perfect score: 178

Sequence: 1 TYNNLLPPRPRIKEVLTIDIFAKVLEKGVDSRS 33

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	173	97.2	956	2 Q8R004	Q8R004 mus musculus
2	173	97.2	987	2 Q99K74	Q99K74 mus musculus
3	173	97.2	989	1 T100 HUMAN	Q75448 h thyroid h
4	169	94.9	956	2 Q9WV1	Q9WV1 mus musculus
5	109	61.2	988	2 Q6P153	Q6P153 xenopus lae
6	109	61.2	988	2 AAH43802	AAH43802 xenopus l
7	99	55.6	984	2 Q6NUW0	Q6NUW0 brachydanio
8	99	55.6	984	2 AAH68408	AAH68408 brachydano
9	61	34.3	412	2 Q9N4T2	Q9N4T2 caenorhabdi
10	59	33.1	1438	2 Q9ZET7	Q9ZET7 mycobacteri
11	59	33.1	5990	2 Q9RLP6	Q9RLP6 mycobacteri
12	57	32.0	314	1 YK67 YEAST	YK673 saccharomyc
13	57	32.0	486	1 RFCL_METMP	Q6M069 methanococc
14	57	32.0	486	2 CAF29878	CAF29878 methanoco
15	57	32.0	603	2 Q8MMB8	Q8MMB8 drosophila
16	56.5	31.7	384	1 VATC ASCSS	Q9ndrs ascidia syd
17	56.5	31.7	7191	2 Q6XA09	Q6XA09 alternaria
18	56.5	31.7	7191	2 AAP78735	AAP78735 alternari
19	55	31.5	262	2 Q28763	Q28763 archaeoglob
20	55.5	31.2	375	2 Q8A079	Q8A079 bacteroides
21	55.5	31.2	515	2 Q8U927	Q8U927 agrobacteri
22	55.5	31.2	531	2 Q7CSU1	Q7CSU1 agrobacteri
23	55.5	31.2	936	2 Q7SM25	Q7SM25 oryza sativ
24	55.5	31.2	936	2 AAS98832	AAS98832 oryza sat
25	55.5	31.2	1351	2 Q6BRL4	Q6BRL4 debaryomyce
26	55.5	31.2	2374	2 Q7S4J9	Q7S4J9 neurospora
27	55	30.9	66	2 Q6ZS35	Q6ZS35 oryza sativ
28	55	30.9	66	2 BAC83465	BAC83465 oryza sat
29	55	30.9	218	2 Q79WY2	Q79WY2 streptococc
30	55	30.9	218	2 Q8K7T6	Q8K7T6 streptococc
31	55	30.9	218	2 Q8PIA2	Q8PIA2 streptococc

32	55	30.9	218	2 Q9A051	Q9A051 streptococc
33	55	30.9	465	2 Q80U04	Q80U04 mus musculus
34	55	30.9	529	1 NR01 SCHPO	Q09702 schizosacch
35	55	30.9	934	2 Q6MY71	Q6MY71 aspergillus
36	55	30.9	934	2 CAF32132	CAF32132 aspergill
37	54.5	30.6	285	2 Q6FJ18	Q6FJ18 candida glia
38	54	30.3	156	2 Q71E11	Q71E11 bacillus su
39	54	30.3	156	2 AAQ08766	AAQ08766 bacillus
40	53.5	30.1	402	2 Q7VKG3	Q7VKG3 haemophilus
41	53	29.8	88	1 Y175_STRR6	Y175_STRR6 streptococ
42	53	29.8	88	1 Y192_STRPN	Y192_STRPN streptococ
43	53	29.8	88	2 Q73HK1	Q73HK1 wolbachia p
44	53	29.8	88	2 AAS14264	AAS14264 wolbachia
45	53	29.8	139	2 Q8UIB4	Q8UIB4 pyrococcus

ALIGNMENTS

RESULT 1
Q8R004
ID Q8R004 PRELIMINARY; PRT; 956 AA.
AC Q8R004;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE KIAA0130-like protein.
GN Name=Thrap4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN (1)_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=ILS, and ISS;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RL Submitted (FEB-2002), to the EMBL/GenBank/DBJ databases.
DR EMBL; AF483498; AAL90772.1; -;
DR MGD; MGI:1344385; Thrap4.
DR GO; GO:0000119; C:mediator complex; IDA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IMP.
SQ SEQUENCE 956 AA; 105941 MW; 41FC6F4D850D6540 CRC64;

Query Match 97.2%; Score 173; DB 2; Length 956;
Best Local Similarity 100.0%; Pred. No. 9.7e-15;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YNNLLPPRPRIKEVLTIDIFAKVLEKGVDSRS 33
DB 664 YNNLLPPRPRIKEVLTIDIFAKVLEKGVDSRS 695

RESULT 2

Q99K74
ID Q99K74 PRELIMINARY; PRT; 987 AA.
AC Q99K74;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Thrap4 protein.
GN Name=Thrap4;
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCB1_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mix FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
 RA Rodriguez A.C., Griewood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RV [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=mix FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC005409; AA05409.1; -;
 DR MGD; MGI:1344385; Thrap4.
 DR GO; GO:0000119; C:mediator complex; IDA.
 DR GO; GO:0006366; P:transcription from Pol II promoter; IMP.
 SQ SEQUENCE 987 AA; 109984 MW; 7136A02C3C0B294 CRC64;
 Query Match 97.2%; Score 173; DB 2; Length 987;
 Best Local Similarity 100.0%; Pred. No. 1e-14;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 YWNLPPKRPKEVLTDFIAKVLKGGWDSRS 33
 DB 695 YWNLPPKRPKEVLTDFIAKVLKGGWDSRS 726
 RESULT 3
 TT100 HUMAN
 ID T100 HUMAN STANDARD; PRT; 989 AA.
 AC 075448; Q14143; Q9NNY5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Thyroid hormone receptor-associated protein complex 100 kDa component
 DE (Trap100) (Thyroid hormone receptor associated protein 4) (Activator-
 DE receptor-interacting protein complex component DRIP100) (Activator-
 DE recruited cofactor 100 kDa component) (ARC100).
 GN Name=THRAP4; Synonyms=TRAP100, DRIP100, ARC100, KIAA0130;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
 RX MEDLINE=98318590; PubMed=9653119;
 RA Yuan C.-X., Ito M., Fondell J.D., Fu Z.-Y., Roeder R.G.;
 RT "The TRAP220 component of a thyroid hormone receptor-associated
 RT protein (TRAP) coactivator complex interacts directly with nuclear
 RT receptors in a ligand-dependent fashion.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:7939-7944 (1998).
 RN [2]
 RP SEQUENCE FROM N.A., AND IDENTIFICATION IN DRIP COMPLEX.
 RX MEDLINE=98301435; PubMed=9637861;
 RA Rachez C., Suldan Z., Ward J., Chang C.-P.B., Burakov D.,
 RA Erdjument-Bromage H., Tempst P., Freedman L.P.;
 RT "A novel protein complex that interacts with the vitamin D3 receptor
 RT in a ligand-dependent manner and enhances VDR transactivation in a
 RT cell-free system.";
 RL Genes Dev. 12:1787-1800(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
 RA Rodriguez A.C., Griewood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RV [5]
 RP IDENTIFICATION IN ARC COMPLEX, AND SEQUENCE OF 1-11 AND 957-965.
 RX MEDLINE=99249346; PubMed=10235267; DOI=10.1038/19789;
 RA Naeae A.M., Beaurang P.A., Zhou S., Abraham S., Solomon W.B.,
 RA Tjian R.;
 RT "Composite co-activator ARC mediates chromatin-directed
 RT transcriptional activation.";
 RL Nature 398:828-832(1999).
 CC -!- FUNCTION: Plays a role in transcriptional coactivation. Plays a
 CC role in nuclear hormone receptor-mediated transactivation.
 CC -!- SUBUNIT: Subunit of the large multiprotein complexes ARC/DRIP and
 CC TRAP. Interacts and co-precipitates with Trap220. Does not
 CC directly bind nuclear hormone receptors.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Abundant in skeletal muscle, heart
 CC and placenta.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF055995; AAC39855.1; -;
 CC EMBL; D50920; BAA09479.2; AUT_INIT.
 CC EMBL; AF277379; AAF78764.1; -;
 CC EMBL; BC011375; AAH11375.1; -;
 CC Genew; HGNC:22963; THRAP4.
 CC MIM; 607000; -;
 CC -----

DR GO:0000119; C-mediator complex; NAS.
DR GO:0005634; C:nucleus; IDA.
DR GO:0003074; F:ligand-dependent nuclear receptor transcrip. . ; NAS.
DR GO:0004872; F:receptor activity; IDA.
DR GO:0016455; F:RNA polymerase II transcription mediator ac. . ; IDA.
DR GO:0046966; F:thyroid hormone receptor binding; IDA.
DR GO:0003712; F:transcription cofactor activity; IDA.
DR GO:0016563; F:transcriptional activator activity; IDA.
DR GO:0042809; F:vitamin D receptor binding; NAS.
DR GO:0030521; F:androgen receptor signaling pathway; IDA.
DR GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR GO:0006367; P:transcription initiation from Pol II promoter; IDA.
KW ATP-binding; Direct protein sequencing; Nuclear protein; Repeat;
FT Transcription regulation; Zinc-finger
FW DOMAIN 128 861
FT REPEAT 128 861
FT FT 1. 6 X REPEATS OF L-X-X-L-L.
FT REPEAT 128 132
FT REPEAT 344 348
FT REPEAT 448 452
FT REPEAT 557 561
FT REPEAT 788 792
FT REPEAT 857 861
FT ZN_FING 93 117
FT NP_BIND 440 447
FT CONFLICT 20 20
FT CONFLICT 204 204
FT CONFLICT 555 555
FT SEQUENCE 989 AA; 110304 MW; CCCE7D4E74D689DC CRC64;
Query Match 97.2%; Score 173; DB 1; Length 989;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 33
DB 697 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 728

RESULT 4
Q9WF1 PRELIMINARY; PRT; 956 AA.

ID Q9WF1
AC Q9WF1; 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE 100 kDa thyroid hormone receptor associated protein.
GN Name=Thrap4; Synonyms=Trap100;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99333122; PubMed=10406464;
RZ Zhang J., Fondall J.D.;
RT "Identification of mouse TRAP100: a transcriptional coregulatory factor for thyroid hormone and vitamin D receptors."
RL Mol. Endocrinol. 13:1130-1140(1999).
DR EMBL; AF126543; AAD42776.1; -.
DR MGD; MGI:1344385; Thrap4.
DR GO:0000119; C-mediator complex; IDA.
DR GO:0006366; P:transcription from Pol II promoter; IMP.
KW Receptor.
FW SEQUENCE 956 AA; 106062 MW; 2DB3B4D9C5E45A4D CRC64;
Query Match 94.9%; Score 169; DB 2; Length 956;
Best Local Similarity 96.9%; Pred. No. 3.5e-14;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 33
DB 664 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 695

RESULT 5
QSP153 PRELIMINARY; PRT; 988 AA.

ID QSP153
AC QSP153;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Thrapp4-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold S.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Hsieh F., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Ra Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shcherchenko V., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smallus D.E., Schnurch A., Schein J.E., Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W., Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043802; AAAH43802.1; -.
DR InterPro; IPR006162; Ppantne.S.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN 1.
SQ SEQUENCE 988 AA; 110513 MW; B780452FE56B03C6 CRC64;
Query Match 61.2%; Score 109; DB 2; Length 988;
Best Local Similarity 65.6%; Pred. No. 6.7e-06;
Matches 21; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 33
DB 695 YIHLPPKTPKEVLRGAFSGSTLEKGVDSRT 726

RESULT 6
AAH43802 PRELIMINARY; PRT; 988 AA.

ID AAH43802
AC AAH43802;
DT 02-MAR-2004 (TRENBLrel. 27, Created)
DT 02-MAR-2004 (TRENBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TRENBLrel. 27, Last annotation update)
DE Thrapp4-prov protein.

GO; GO:000119; C:mediator complex; NAS.
GO; GO:0005634; C:nucleus; IDA.
GO; GO:000374; F:ligand-dependent nuclear receptor transcrip. . ; NAS.
GO; GO:0004872; F:receptor activity; IDA.
GO; GO:0016455; F:RNA polymerase II transcription mediator ac. . ; IDA.
GO; GO:0046966; F:thyroid hormone receptor binding; IDA.
GO; GO:0003712; F:transcription cofactor activity; IDA.
GO; GO:0016563; F:transcriptional activator activity; IDA.
GO; GO:0042809; F:vitamin D receptor binding; NAS.
GO; GO:0030521; F:androgen receptor signaling pathway; IDA.
GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
GO; GO:0006367; P:transcription initiation from Pol II promoter; IDA.
KW ATP-binding; Direct protein sequencing; Nuclear protein; Repeat;
Transcription regulation; Zinc-finger
FT DOMAIN 128 861
FT REPEAT 128 861
FT FT 1. 6 X REPEATS OF L-X-X-L-L.
FT REPEAT 344 348
FT REPEAT 448 452
FT REPEAT 557 561
FT REPEAT 788 792
FT REPEAT 857 861
FT ZN_FING 93 117
FT NP_BIND 440 447
FT CONFLICT 20 20
FT CONFLICT 204 204
FT CONFLICT 555 555
FT SEQUENCE 989 AA; 110304 MW; CCCEDE7D4E74D689DC CRC64;
Query Match 97.2%; Score 173; DB 1; Length 989;
Best Local Similarity 100.0%; Pred. No. 1e-14;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 33
DB 697 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 728

RESULT 4
Q9WF1 PRELIMINARY; PRT; 956 AA.
AC Q9WF1
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 100 kDa thyroid hormone receptor associated protein.
GN Name=Thrap4; Synonyms=Trap100;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99333122; PubMed=10406464;
RZ Zhang J., Fondall J.D.;
RT "Identification of mouse TRAP100: a transcriptional coregulatory factor for thyroid hormone and vitamin D receptors."
RL Mol. Endocrinol. 13:1130-1140(1999).
DR EMBL; AF126543; AAD42776.1; -.
DR MGD; MGI:1344385; Thrap4.
DR GO; GO:000119; C:mediator complex; IDA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IMP.
KW Receptor.
SQ SEQUENCE 956 AA; 106062 MW; 2DB3B4D9C5E45A4D CRC64;
Query Match 94.9%; Score 169; DB 2; Length 956;
Best Local Similarity 96.9%; Pred. No. 3.5e-14;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 33
DB 664 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 695

RESULT 5
QSP153 PRELIMINARY; PRT; 988 AA.
AC QSP153
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Thrsp4-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold S.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Hsieh F., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Ra Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shcherchenko V., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smallus D.E., Schnurch A., Schein J.E., Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W., Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043802; AAAH43802.1; -.
DR InterPro; IPR006162; Ppantne.S.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN 1.
SQ SEQUENCE 988 AA; 110513 MW; B780452FE56B03C6 CRC64;
Query Match 61.2%; Score 109; DB 2; Length 988;
Best Local Similarity 65.6%; Pred. No. 6.7e-06;
Matches 21; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 YWNLLPPKPIKEVLTIDIFAKVLEKGWDSRS 33
DB 695 YIHLPPKTIKEVLRGAFSGSTLEKGVDSRT 726

RESULT 6
AAH43802 PRELIMINARY; PRT; 988 AA.
ID AAH43802
AC AAH43802
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Thrsp4-prov protein.

OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA MEDLINE=223441132; PubMed=12454917;
 RT Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative";
 RL Dev. Dyn. 225:384-391 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klein S.L., Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Kettman M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RA and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Strausberg R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC043802; AAH43802.1; -;
 SQ SEQUENCE 988 AA; 110513 MW; B780452FE368D3C6 CRC64;
 Query Match 61.2%; Score 109; DB 2; Length 988;
 Best Local Similarity 65.6%; Pred. No. 6,7e-06;
 Matches 21; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
 QY 2 YWNLPPKPIKEVLTDFIAKVLKGVDSRS 33
 DB 695 YIHLPPKPIKEVLTDFIAKVLKGVDSRS 726
 RESULT 7
 Q6NUWO PRELIMINARY; PRT; 984 AA.
 ID Q6NUWO
 AC Q6NUWO;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DE Hypothetical protein (Fragment).
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klein S.L., Strausberg R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC043802; AAH43802.1; -;
 SQ SEQUENCE 988 AA; 110513 MW; B780452FE368D3C6 CRC64;
 Query Match 61.2%; Score 109; DB 2; Length 988;
 Best Local Similarity 65.6%; Pred. No. 6,7e-06;
 Matches 21; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
 QY 2 YWNLPPKPIKEVLTDFIAKVLKGVDSRS 33
 DB 695 YIHLPPKPIKEVLTDFIAKVLKGVDSRS 726

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Kettman M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RA and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Strausberg R.;
 RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC068408; AAH68408.1; -;
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 984 AA; 110289 MW; 9567D437043F536D CRC64;
 Query Match 55.6%; Score 99; DB 2; Length 984;
 Best Local Similarity 59.4%; Pred. No. 0.00016;
 Matches 19; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
 QY 2 YWNLPPKPIKEVLTDFIAKVLKGVDSRS 33
 DB 691 YNLLPAPRLPQALQSQFRKVLKGVDSHA 722
 RESULT 8
 AAH68408 PRELIMINARY; PRT; 984 AA.
 ID AAH68408
 AC AAH68408;
 DT 14-APR-2004 (TrEMBLrel. 27, Created)
 DT 14-APR-2004 (TrEMBLrel. 27, Last sequence update)
 DE Hypothetical protein (Fragment).
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Kettman M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RA and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Strausberg R.;
 RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC068408; AAH68408.1; -;
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 984 AA; 110289 MW; 9567D437043F536D CRC64;
 Query Match 55.6%; Score 99; DB 2; Length 984;
 Best Local Similarity 59.4%; Pred. No. 0.00016;
 Matches 19; Conservative 4; Mismatches 9; Indels 0; Gaps 0;
 QY 2 YWNLPPKPIKEVLTDFIAKVLKGVDSRS 33
 DB 691 YNLLPAPRLPQALQSQFRKVLKGVDSHA 722

RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
[g]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC06795; AAF59493.2; -;
DR HSP; P11362; 1FGK.
DR WormPep; Y5OD4B.6; CE261136.
DR GO; GO:0005524; F.ATP binding; IEA.
DR GO; GO:0004713; F.protein-tyrosine kinase activity; IEA.
DR GO; GO:0006468; P.protein amino acid phosphorylation; IEA.
DR InterPro; IPR011009; Kinase-like.
DR InterPro; IPR000719; Prot.kinase.
DR InterPro; IPR01245; Tyr_pkinase.
DR InterPro; IPR008266; Tyr_pkinase_AS.
DR Pfam; PF00069; Pkinase; I.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SMO0219; TyrcK; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; UNKNOWN 1.
KW Hypothetical protein; Kinase; transferase; tyrosine-protein kin
SQ SEQUENCE 412 AA; 48199 MW; 8E5C389FFA726659 CRG64;

Query Match 34.3%; Score 61; DB 2; Length 412;
Best Local Similarity 45.8%; Pred.No.11;
Matches 11; Conservative 4; Mismatches 9; Indels 0;

QY 3 WNLLPKRPIKEVLTDIFAKVLEK 26
||| ||| | : : : : :
DB 376 WNLNPKRVFSECRDFQKLQQ 399

RESULT 10
Q9ZET7
ID Q9ZET7 PRELIMINARY; PRT; 1438 AA.
AC Q9ZET7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Peptide synthetase (fragment).
GN Name=peptide synthetase;
GS Mycobacterium smegmatis;
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mc2155;
RA Billman-Jacobs H., McConville M., Haites R., Kovacevic S.;
RL Submitted (OCI-1998) to the EMBL/GenBank/DDBJ databases.
CC - SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme family.
CC
CC EMBL; AJ011839; CAA09803.1; -;
DR GO; GO:0003824; F.catalytic activity; IEA.
DR GO; GO:0048037; F.co-factor binding; IEA.
DR GO; GO:0008152; P.metabolism; IEA.
DR InterPro; IPR010071; AA_adenyl_dom.
DR InterPro; IPR009081; ACP-like.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR00162; Ppartne.S.
DR InterPro; IPR006163; Pp_bind".
DR Pfam; PF00501; AMP-binding; 1.
DR Pfam; PF00668; Condensation; 2.
DR Pfam; PF00550; PP-binding; 1.
DR PRINTS; PR00154; AMPBINDING.
DR TIGRFAMS; TIGR01733; AA-adenyl-dom; 1.
DR PROSITE; PS50075; ACP DOMAIN; 1.

```

DR PROSITE: PS00455; AMP BINDING; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
FT NON_TER 1438 1438
SQ SEQUENCE 1438 AA; 153614 MW; C90E52B84F2DA40E CRC64;

Query Match 33.1%; Score 59; DB 2; Length 1438;
Best Local Similarity 52.4%; Pred. No. 78;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 NLLPPKRIPEVLTIDIFAKVL 24
DB 946 NYRAPSNPVEEVLADIYAQVL 966

RESULT 11
Q9RLP6 PRELIMINARY; PRT; 5990 AA.
AC Q9RLP6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Peptide synthetase.
GN Name=mps;
OS Mycobacterium smegmatis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC;
RA MEDLINE=99440174; PubMed=10510238;
RA Billman-Jacobs H., McConville M.C., Haites R.E., Kovacevic S.,
Coppel R.L.;
RT "Identification of a peptide synthetase involved in the biosynthesis
of glycopeptidolipids of Mycobacterium smegmatis.";
RL Mol. Microbiol. 33:1244-1253(1999).
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC EMBL; AJ238027; CAB55600.1; -.
DR HSSP; 030409; IDNY.
DR GO; GO:0003824; P: catalytic activity; IEA.
DR GO; GO:0048037; P: cofactor binding; IEA.
DR GO; GO:0008152; P: metabolism; IEA.
DR InterPro; IPR010071; AA-adenyl_dom.
DR InterPro; IPR009081; ACP-like.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR010060; NRPS synth.
DR InterPro; IPR006162; Peptide S.
DR InterPro; IPR006163; Peptide S.
DR InterPro; IPR010080; Thioester_redct.
DR Pfam; PF00501; AMP-binding; 4.
DR Pfam; PF00668; Condensation; 7.
DR Pfam; PF00550; PP-binding; 4.
DR PRINTS; PR00154; AMPBINDING.
DR TIGRFAMs; TIGR01733; AA-adenyl-dom; 4.
DR TIGRFAMs; TIGR01720; NRPS-para261; 3.
DR TIGRFAMs; TIGR01746; Thioester_redct; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
DR PROSITE; PS00455; AMP BINDING; 4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_4.
SQ SEQUENCE 5990 AA; 645938 MW; 25BA1C934EE902D5 CRC64;

Query Match 33.1%; Score 59; DB 2; Length 5990;
Best Local Similarity 52.4%; Pred. No. 3.6e+02;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 NLLPPKRIPEVLTIDIFAKVL 24
DB 946 NYRAPSNPVEEVLADIYAQVL 966

RESULT 12
YK67 YEAST STANDARD; PRT; 314 AA.
AC P36163;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Hypothetical 35.8 kDa protein in PRP16-SRP40 intergenic region.
GN OrderedLocusNames=YKR087C; ORFNames=YKR407;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94262327; PubMed=8203164;
RX Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
RT "The complete sequence of an 18,002 bp segment of Saccharomyces
cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,
and six new open reading frames.";
RL Yeast 10:231-245(1994).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; Z27116; CAA81638.1; -.
DR EMBL; Z28312; CAA82166.1; -.
DR PIR; S38165; S38165.
DR GeneOnline; 140066; -.
DR MEROPS; M48_018; -.
DR SGD; S0001795; YKR087C.
DR GO; GO:0005743; C: mitochondrial inner membrane; IEA.
DR GO; GO:0005739; C: mitochondrion; IEA.
DR GO; GO:0004222; P: metalloendopeptidase activity; IMP.
DR GO; GO:0006515; P: misfolded or incompletely synthesized prote. .; IMP.
DR InterPro; IPR001915; Peptidase M48.
DR Pfam; PF01435; Peptidase M48; 1.
DR Hypothetical protein; Transmembrane.
FT TRANSMEM 36 52 Potential.
FT TRANSMEM 142 161 Potential.
FT TRANSMEM 191 208 Potential.
SQ SEQUENCE 314 AA; 35846 MW; 22DD7EE3C747A440 CRC64;

Query Match 32.0%; Score 57; DB 1; Length 314;
Best Local Similarity 33.3%; Pred. No. 28;
Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 TYNNLLPPKRIPEVLTIDIFAKVLEKGVWD 30
DB 86 TQOEILPQHPPLSIKINIFMKIVEAYKD 115

RESULT 13
RFCL_METMP STANDARD; PRT; 486 AA.
AC Q6M0E9;
DT 01-OCT-2004 (Rel. 45, Created)
DT 01-OCT-2004 (Rel. 45, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Replication factor C large subunit (RFC large subunit) (Clamp loader
large subunit).
GN Name=rfcl; Synonyms=rfcb; OrderedLocusNames=MWP0322;
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.

```

Matches	10;	Conservative	4;	Mismatches	6;	Indels	0;	Gaps
---------	-----	--------------	----	------------	----	--------	----	------

QY	7	PPGPIKEVLTDIFAKVLEK 26
	: :	: :
Db	446	PPKZPLKEVIETLEKSVEK 465


```

RESULT 15
Q8NMWB PRELIMINARY; PRT; 603 AA.
AC Q8MMB8;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-WAR-2003 (TReMBLrel. 23, Last annotation update)
DE CG30125-PA.
GN ORFNames=CG30125;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SE SEQUENCE FROM N.A.
RP MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yeandle M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Fiankoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., CWiley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush C., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kratt C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Mazzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J., Ye J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zavert J.S., Zhan M., Zhang G., Zhao Q., Zhou L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
SE SEQUENCE FROM N.A.
RP MEDLINE=22426065; PubMed=12537568;
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzy D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Swirskas R., Taber F.E., Wan X., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
    melanogaster euchromatic genome sequence."

```

```

RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celisner S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.D., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FLYBASE;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FLYBASE;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003797; AAM68433.1; -
DR FLYBASE; FBgn050125; CG30125.
SQ SEQUENCE 603 AA; 69506 MW; 6FA4805764D3139D CRC64;

```

```

Query Match      32.0%; Score 57; DB 2; Length 603;
Best Local Similarity 35.1%; Pred. No. 57;
Matches 13; Conservative 6; Mismatches 10; Indels 8; Gaps 2;

```

```

Oy 3 WNL----LPPKRPKEVLTDFAKVLEKG----WVDS 31
   | | | | | | | | | | | | | | | | | |
Db 502 WRILGIPLNPKSPURFMFKDYFYRLSGRLREQWHS 538

```

Search completed: November 10, 2004, 12:27:15
Job time : 39.1065 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 5.65995 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-17
Perfect score: 101
Sequence: 1 LFTLLTLWTMRCSTSPSG 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PGTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	43.6	175	4	US-09-252-991A-20024
2	44	43.6	532	3	US-08-911-321-11
3	44	43.6	532	4	US-08-504-617-3
4	44	43.6	532	5	PCT-US95-13975-2
5	43	42.6	172	4	US-09-252-991A-23710
6	43	42.6	183	4	US-09-270-767-31682
7	42	41.6	94	4	US-09-248-796A-26505
8	42	41.6	222	4	US-09-148-545-135
9	42	41.6	229	4	US-09-252-991A-27823
10	42	41.6	248	4	US-09-252-991A-32999
11	42	41.6	436	4	US-09-252-991A-30911
12	41.5	41.1	166	4	US-09-621-976-5265
13	41.5	41.1	266	4	US-08-311-731A-112
14	41	40.6	173	4	US-09-252-991A-30903
15	41	40.6	251	4	US-09-149-476-568
16	41	40.6	239	3	US-09-147-826B-2
17	41	40.6	326	1	US-08-190-802A-39
18	41	40.6	326	3	US-08-477-346-39
19	41	40.6	326	3	US-08-473-089-39
20	41	40.6	326	4	US-08-487-072A-39
21	41	40.6	340	1	US-08-190-802A-38
22	41	40.6	340	1	US-08-190-802A-40
23	41	40.6	340	1	US-08-190-802A-42
24	41	40.6	340	3	US-09-180-783-2
25	41	40.6	340	3	US-08-477-346-38
26	41	40.6	340	3	US-08-477-346-40
27	41	40.6	340	3	US-08-477-346-42

28	41	40.6	340	3	US-08-473-089-38
29	41	40.6	340	3	US-08-473-089-40
30	41	40.6	340	3	US-08-473-089-42
31	41	40.6	340	4	US-08-487-072A-38
32	41	40.6	340	4	US-08-487-072A-40
33	41	40.6	340	4	US-08-487-072A-42
34	41	40.6	340	4	US-09-245-039-1
35	41	40.6	340	4	US-09-538-092-848
36	41	40.6	340	4	US-09-538-092-899
37	41	40.6	340	4	US-09-538-092-940
38	41	40.6	341	1	US-08-190-802A-45
39	41	40.6	341	3	US-08-477-346-45
40	41	40.6	341	3	US-08-473-089-45
41	41	40.6	341	4	US-08-487-072A-45
42	41	40.6	359	4	US-09-252-991A-29838
43	41	40.6	424	4	US-09-489-039A-7646
44	41	40.6	456	2	US-08-910-731-2
45	41	40.6	456	2	US-08-795-395-2

ALIGNMENTS

RESULT 1
US-09-252-991A-20024
; Sequence 20024, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20024
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20024

Query Match 43.6%; Score 44; DB 4; Length 175;
Best Local Similarity 53.8%; Pred No 16;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 TLWTMRCSTSPSG 19
DB 44 TNWTTGCAGTPTG 56

RESULT 2
US-08-911-321-11
; Sequence 11, Application US/08911321
; Patent No. 6010703
; GENERAL INFORMATION:
; APPLICANT: Roger K. Maes and Stephen J. Spatz
; TITLE OF INVENTION: Recombinant Poxvirus
; TITLE OF INVENTION: Vaccine Against
; TITLE OF INVENTION: Feline Rhinotracheitis
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb

us-10-092-750-17.ra1

Fri Nov 12 14:54:53 2004

```

;
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,321
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/096,183
; FILING DATE: July 26, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-166
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; TELEX: No. 6010703e
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Polypeptide
; DESCRIPTION: No
; HYPOTHEICAL: No
; ANTI-SENSE: No
; ORIGINAL SOURCE:
; ORGANISM: Feline herpesvirus-1
; STRAIN: 1
; INDIVIDUAL ISOLATE: C-27
; CELL TYPE: N/A
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD: Deduced Sequence
; OTHER INFORMATION: 9B
;
; US-08-911-321-11
;
; Query Match 43.6%; Score 44; DB 3; Length 532;
; Best Local Similarity 70.0%; Pred. No. 54;
; Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
;
; QY 9 WTMRCSSTPS 18
; DB 290 WTSRCINTPS 299
;
; RESULT 3
; PCT-US95-617-3
; Sequence 3, Application US/08504617
; Patent No. 6521236
; GENERAL INFORMATION:
; APPLICANT: Willemse, Martha Jacoba
; APPLICANT: Sondermeijer, Paulus Jacobus Antonius
; TITLE OF INVENTION: Vector vaccines of recombinant
; TITLE OF INVENTION: Feline herpesviruses
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 6521236el
; STREET: 1330 Ficcard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13975
; FILING DATE: 26-OCT-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/329,883
; FILING DATE: 26-OCT-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39118-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-13975-2
;
; APPLICATION NUMBER: US/08/504,617
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/211,150
; FILING DATE: March 22, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gortley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; OTHER INFORMATION: /label= ORF-2
;
; US-08-504-617-3
;
; Query Match 43.6%; Score 44; DB 4; Length 532;
; Best Local Similarity 70.0%; Pred. No. 54;
; Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
;
; QY 9 WTMRCSSTPS 18
; DB 290 WTSRCINTPS 299
;
; RESULT 4
; PCT-US95-13975-2
; Sequence 2, Application PC/TUS9513975
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: McDonnell, Michael W.
; TITLE OF INVENTION: Recombinant Feline Herpes virus
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13975
; FILING DATE: 26-OCT-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/329,883
; FILING DATE: 26-OCT-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39118-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0400
; TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-13975-2

```

```

Query Match      43.6%; Score 44; DB 5; Length 532;
Best Local Similarity 70.0%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 9 WTMRCSTPS 18
Db 290 WTSRCINTPS 299

RESULT 5
US-09-252-991A-23710
; Sequence 23710, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23710
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23710

Query Match      42.6%; Score 43; DB 4; Length 172;
Best Local Similarity 63.6%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 WTMRCSTPSG 19
Db 6 WRRRCASWPSG 16

RESULT 6
US-09-270-767-31682
; Sequence 31682, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62617
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 31682
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-31682

Query Match      42.6%; Score 43; DB 4; Length 183;
Best Local Similarity 53.8%; Pred. No. 25;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 7 TLWTMRCSTSPSG 19
Db 77 TTWIMACAYAPSG 89

RESULT 7
US-09-248-796A-26505
; Sequence 26505, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al

```

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN;
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 26505
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-26505

Query Match      41.6%; Score 42; DB 4; Length 94;
Best Local Similarity 52.9%; Pred. No. 18;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 LFTILLTLWTMRCSTP 17
Db 33 LTTLLHLFTINMSTTP 49

RESULT 8
US-09-148-545-135
; Sequence 135, Application US/09148545
; Patent No. 6590075
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: PZ001P1
; CURRENT APPLICATION NUMBER: US/09/148,545
; CURRENT FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: PCT/US98/04482
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,161
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23

```

	EARLIER APPLICATION NUMBER:	60/047,581	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,879	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,584	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,880	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,500	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,894	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,587	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,911	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,492	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,836	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,598	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,874	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,613	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,910	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,582	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,864	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,596	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,612	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,845	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,632	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/056,892	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,601	EARLIER FILING DATE:	1997-08-22
	EARLIER APPLICATION NUMBER:	60/047,595	EARLIER FILING DATE:	1997-05-23
	EARLIER APPLICATION NUMBER:	60/043,580	EARLIER FILING DATE:	1997-05-23
	EARLIER APPLICATION NUMBER:	60/057,761	EARLIER FILING DATE:	05-Sep-1997
	EARLIER APPLICATION NUMBER:	60/043,568	EARLIER APPLICATION NUMBER:	60/047,599
	EARLIER APPLICATION NUMBER:	60/043,314	EARLIER APPLICATION NUMBER:	60/047,588
	EARLIER APPLICATION NUMBER:	60/043,569	EARLIER APPLICATION NUMBER:	60/047,585
	EARLIER APPLICATION NUMBER:	60/043,311	EARLIER APPLICATION NUMBER:	60/047,586
	EARLIER APPLICATION NUMBER:	60/043,671	EARLIER APPLICATION NUMBER:	60/047,590
	EARLIER APPLICATION NUMBER:	60/043,674	EARLIER APPLICATION NUMBER:	60/047,594
	EARLIER APPLICATION NUMBER:	60/043,669	EARLIER APPLICATION NUMBER:	60/047,589
	EARLIER APPLICATION NUMBER:	60/043,312	EARLIER APPLICATION NUMBER:	60/047,593
	EARLIER APPLICATION NUMBER:	60/043,313	EARLIER APPLICATION NUMBER:	60/047,614
	EARLIER APPLICATION NUMBER:	60/043,672	EARLIER APPLICATION NUMBER:	60/043,578
	EARLIER APPLICATION NUMBER:	60/043,315	EARLIER APPLICATION NUMBER:	60/043,576
	EARLIER APPLICATION NUMBER:	60/048,974	EARLIER APPLICATION NUMBER:	60/047,501
	EARLIER APPLICATION NUMBER:	60/056,886	EARLIER APPLICATION NUMBER:	60/043,670
	EARLIER APPLICATION NUMBER:	60/056,877	EARLIER APPLICATION NUMBER:	60/056,632
	EARLIER APPLICATION NUMBER:	60/056,889	EARLIER APPLICATION NUMBER:	60/056,664
	EARLIER APPLICATION NUMBER:	60/056,893	EARLIER APPLICATION NUMBER:	60/056,876
	EARLIER APPLICATION NUMBER:	60/056,630	EARLIER APPLICATION NUMBER:	60/056,881
	EARLIER APPLICATION NUMBER:	60/056,878	EARLIER APPLICATION NUMBER:	60/056,909
	EARLIER APPLICATION NUMBER:	60/056,662	EARLIER APPLICATION NUMBER:	60/056,875
	EARLIER APPLICATION NUMBER:	60/056,872	EARLIER APPLICATION NUMBER:	60/056,862
	EARLIER APPLICATION NUMBER:	60/056,882	EARLIER APPLICATION NUMBER:	60/056,887
	EARLIER APPLICATION NUMBER:	60/056,637	EARLIER APPLICATION NUMBER:	60/056,908
	EARLIER APPLICATION NUMBER:	60/056,903	EARLIER APPLICATION NUMBER:	60/048,964
	EARLIER APPLICATION NUMBER:	60/056,888	EARLIER APPLICATION NUMBER:	1997-06-06

```

; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 135
; LENGTH: 222

Query Match      41.6%; Score 42; DB 4; Length 222;
Best Local Similarity 77.8%; Pred. No. 44;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 LTLWTMRC 14
   |||:||||
Db 142 LDLWTMRC 150

RESULT 9
US-09-252-991A-27823
; Sequence 27823, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27823
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27823

Query Match      41.6%; Score 42; DB 4; Length 229;
Best Local Similarity 60.0%; Pred. No. 46;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 WTMRCSTPS 18
   |||:||||
Db 36 WSTRCAGTSP 45

RESULT 10
US-09-252-991A-32999
; Sequence 32999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32999
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32999

Query Match      41.6%; Score 42; DB 4; Length 248;
Best Local Similarity 70.0%; Pred. No. 50;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 10 TMRCSSTPSG 19
   |||:||||
Db 41 SMRCASSTPPG 50

RESULT 11
US-09-252-991A-30911
; Sequence 30911, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30911
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30911

Query Match      41.6%; Score 42; DB 4; Length 436;
Best Local Similarity 75.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 LTLWTMRC 13
   |||:||||
Db 158 LTMWTIRC 165

RESULT 12
US-09-621-976-5265
; Sequence 5265, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5265
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -21...-1
US-09-621-976-5265

Query Match      41.1%; Score 41.5; DB 4; Length 116;
Best Local Similarity 50.0%; Pred. No. 27;
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 2 FTLLTLWTMRCSTP 17
   |||:||||
Db 9 FTMLLAMLV-CGSEP 23

RESULT 13
US-08-311-731A-112
; Sequence 112, Application US/08311731A

```

Patent No. 6583266
 GENERAL INFORMATION:
 APPLICANT: SMITH, DOUGLAS
 APPLICANT: NAO, JEN-I
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
 TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 411
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 STREET: 600 ATLANTIC AVENUE
 CITY: BOSTON
 STATE: MASSACHUSETTS
 COUNTRY: USA
 ZIP: 02210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA: US/08/311,731A
 APPLICATION NUMBER: US/08/311,731A
 FILING DATE:
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: GATES, EDWARD R.
 REGISTRATION NUMBER: 31,616
 REFERENCE/DOCKET NUMBER: C0044/7125
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/720-3500
 TELEFAX: 617/720-2441
 INFORMATION FOR SEQ ID NO: 112:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 266 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: MYCOBACTERIUM LEPRAE
 US-08-311-731A-112

Query Match 41.1%; Score 41.5; DB 4; Length 266;
 Best Local Similarity 32.0%; Pred. No. 65;
 Matches 8; Conservative 5; Mismatches 3; Indels 9; Gaps 1;

Qy 4 ILLTLWT-----MECSSTPSG 19
 Db 15 LALTAKSKAQSPFFARLEKCTAPAG 39

RESULT 14
 US-09-252-991A-30903
 Sequence 30903, Application us/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 30903
 LENGTH: 173
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-30903

Query Match 40.6%; Score 41; DB 4; Length 173;
 Best Local Similarity 60.0%; Pred. No. 49;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Qy 9 WTMRCSTPS 18
 Db 61 NSWRCSPTPA 70
 RESULT 15
 US-09-149-476-568
 Sequence 568, Application US/09149476
 Patent No. 6420526
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: 186 Human Secreted proteins
 FILE REFERENCE: PZ002P1
 CURRENT APPLICATION NUMBER: US/09/149,476
 CURRENT FILING DATE: 1998-09-08
 EARLIER APPLICATION NUMBER: PCT/US98/04493
 EARLIER FILING DATE: 1998-03-06
 EARLIER APPLICATION NUMBER: 60/040,162
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,333
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/038,621
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,626
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,334
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,336
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,163
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/047,600
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,615
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,597
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,502
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,633
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,583
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,617
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,618
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,503
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,592
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,581
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,584
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,500
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,587
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,492
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,598
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,613
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,582
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,596
 EARLIER FILING DATE: 1997-05-23

; EARLIER APPLICATION NUMBER: 60/047,612
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,632
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,601
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/043,580
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,568
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,314
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,569
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,311
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,671
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,674
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,669
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,312
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,313
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,672
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,315
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/048,974
 ; EARLIER FILING DATE: 1997-08-06
 ; EARLIER APPLICATION NUMBER: 60/056,886
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,877
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,889
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,893
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,630
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,878
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,662
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,872
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,882
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,637
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,903
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,888
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,879
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,880
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,894
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,911
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,864
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,631

; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,845
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,892
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/057,761
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/047,595
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,599
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,588
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,585
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,586
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,590
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,594
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,589
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,593
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,614
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/043,578
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,576
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/047,501
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/043,670
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/056,632
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,664
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,876
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,881
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,909
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,875
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,862
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,887
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,908
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/048,964
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/057,650
 ; EARLIER FILING DATE: 1997-09-05
 ; EARLIER APPLICATION NUMBER: 60/056,884
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/057,669
 ; EARLIER FILING DATE: 1997-09-05
 ; EARLIER APPLICATION NUMBER: 60/049,610
 ; EARLIER FILING DATE: 1997-06-13
 ; EARLIER APPLICATION NUMBER: 60/061,060
 ; EARLIER FILING DATE: 1997-10-02

Query Match 40.6%; Score 41; DB 4; Length 251;
 Best Local Similarity 53.8%; Pred. No. 73;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 LLTWTMRCSTP 17
 : : : : :
 Db 193 IMSLWTQSHSTP 205

Fri Nov 12 14:54:53 2004

us-10-092-750-17.rai

Page 8

Search completed: November 10, 2004, 12:32:30
Job time : 6.65895 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 15:53:52 ; Search time 18.3256 Seconds
(without alignments)
366.225 Million cell updates/sec

Title: US-10-092-750-17
Perfect score: 101
Sequence: 1 LFTILLTLWTMCSTPSG 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	100.0	19	14	US-10-092-750-17
2	52	51.5	151	17	US-10-425-115-240916
3	48	47.5	114	13	US-10-001-870-184
4	47.5	47.0	242	15	US-10-424-599-198283
5	47.5	47.0	542	14	US-10-163-198-25
6	47	46.5	116	14	US-10-231-417-251
7	47	46.5	252	14	US-10-156-761-10847
8	46	45.5	73	17	US-10-425-115-213984
9	46	45.5	80	10	US-09-764-891-3368
10	46	45.5	98	16	US-10-437-963-121796
11	46	45.5	1478	16	US-10-437-963-115917
12	45	44.6	55	15	US-10-424-599-181785
13	45	44.6	84	16	US-10-437-963-127870

14	45	44.6	297	16	US-10-437-963-166688
15	45	44.6	1005	10	US-09-934-070-8
16	45	44.6	1005	15	US-10-222-772-8
17	44	43.6	57	15	US-10-424-599-242110
18	44	43.6	62	15	US-10-424-599-152966
19	44	43.6	115	15	US-10-424-599-184646
20	44	43.6	133	17	US-10-425-115-236842
21	44	43.6	174	17	US-10-425-115-318411
22	44	43.6	310	17	US-10-425-115-318412
23	44	43.6	1077	16	US-10-437-963-113282
24	43.5	43.1	33	9	US-09-864-761-46206
25	43	42.6	52	14	US-10-231-417-193
26	43	42.6	127	17	US-10-425-115-354067
27	43	42.6	518	17	US-10-739-930-6764
28	43	42.6	1922	17	US-10-425-115-297551
29	42.5	42.1	88	17	US-10-425-115-190789
30	42.5	42.1	230	17	US-10-425-115-308281
31	42.5	42.1	231	15	US-10-425-114-49330
32	42	41.6	18	9	US-09-864-761-39272
33	42	41.6	35	9	US-09-864-761-45115
34	42	41.6	46	9	US-09-864-761-35341
35	42	41.6	70	15	US-10-424-599-265206
36	42	41.6	90	17	US-10-425-115-229759
37	42	41.6	113	16	US-10-437-963-106069
38	42	41.6	113	17	US-10-425-115-233167
39	42	41.6	140	15	US-10-425-114-57552
40	42	41.6	168	14	US-10-029-386-33431
41	42	41.6	199	17	US-10-425-115-318589
42	42	41.6	205	14	US-10-029-386-33930
43	42	41.6	213	16	US-10-767-701-54209
44	42	41.6	220	15	US-10-240-240A-14
45	42	41.6	221	9	US-09-799-777-68

ALIGNMENTS

RESULT 1
US-10-092-750-17
; Sequence 17, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 19
; TYPE: PPT
; ORGANISM: Homo sapiens
US-10-092-750-17

Query Match 100.0%; Score 101; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 5.6e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LFTILLTLWTMCSTPSG 19
DB 1 LFTILLTLWTMCSTPSG 19
RESULT 2
US-10-425-115-240916
; Sequence 240916, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 240916
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: MMT4577_151294C.1.pap
 US-10-425-115-240916

Query Match 51.5%; Score 52; DB 17; Length 151;
 Best Local Similarity 66.7%; Pred. No. 7.5;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 LWTMRCSTSPG 19
 DB 92 LWALRCRSPG 103

RESULT 3
 US-10-001-870-184
 ; Sequence 184, Application US/10001870
 ; Publication No. US20020150924A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Salceda, Susana
 ; APPLICANT: Macina, Roberto
 ; APPLICANT: Recipon, Herve
 ; APPLICANT: Sun, Yongming
 ; APPLICANT: Liu, Chenghua
 ; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and
 ; FILE REFERENCE: DEX-0283
 ; CURRENT APPLICATION NUMBER: US/10/001,870
 ; CURRENT FILING DATE: 2001-11-20
 ; PRIOR APPLICATION NUMBER: 60/252,189
 ; PRIOR FILING DATE: 2000-11-21
 ; NUMBER OF SEQ ID NOS: 217
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 184
 ; LENGTH: 114
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 US-10-001-870-184
 Query Match 47.5%; Score 48; DB 13; Length 114;
 Best Local Similarity 47.4%; Pred. No. 22;
 Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
 QY 1 LFTILLTWTMRCSTSPG 19
 DB 48 LFCVVSRPWCNCLSTSPG 66
 RESULT 4
 US-10-424-599-198283
 ; Sequence 198283, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 198283
 ; LENGTH: 242
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MBT3847_21074C.1.pap
 US-10-424-599-198283

Query Match 47.0%; Score 47.5; DB 15; Length 242;
 Best Local Similarity 50.0%; Pred. No. 55;
 Matches 11; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY 1 LFTILLTLW---TMRCSSTSPG 19
 DB 26 LTTIMVTLMWMTFFCWHPGG 47

RESULT 5
 US-10-163-198-25
 ; Sequence 25, Application US/10163198
 ; Publication No. US20030126645A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rebecca E. Cahoon
 ; APPLICANT: Elmer P. Heppard
 ; APPLICANT: No. US20030126645A1uhiro Nagasawa
 ; APPLICANT: Hajime Sakai
 ; TITLE OF INVENTION: Alteration Of Embryo/Endosperm Size During Seed Development
 ; FILE REFERENCE: BBI487 US NA
 ; CURRENT APPLICATION NUMBER: US/10/163,198
 ; CURRENT FILING DATE: 2002-06-05
 ; PRIOR APPLICATION NUMBER: 60/295,921
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: 60/334,317
 ; PRIOR FILING DATE: 2001-11-28
 ; NUMBER OF SEQ ID NOS: 111
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 25
 ; LENGTH: 542
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 US-10-163-198-25

Query Match 47.0%; Score 47.5; DB 14; Length 542;
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;
 Matches 11; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY 1 LFTILLTLW---TMRCSSTSPG 19
 DB 26 LTTIMVTLMWMTFFCWHPGG 47

RESULT 6
 US-10-231-417-251
 ; Sequence 251, Application US/10231417
 ; Publication No. US20030176681A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Feng et al.
 ; TITLE OF INVENTION: 148 Human Secreted Proteins
 ; FILE REFERENCE: PZ019P1
 ; CURRENT APPLICATION NUMBER: US/10/231,417
 ; CURRENT FILING DATE: 2002-08-30
 ; PRIOR APPLICATION NUMBER: US/09/296,622
 ; PRIOR FILING DATE: 1999-04-23
 ; NUMBER OF SEQ ID NOS: 619
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 251
 ; LENGTH: 116
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE

```

; LOCATION: (116)
; OTHER INFORMATION: Xaa equals stop translation
US-10-231-417-251

```

Query Match 46.5%; Score 47; DB 14; Length 116;
Best Local Similarity 52.4%; Pred. No. 32;
Matches 11; Conservative 4; Mismatches 4; Indels

Qy 1 LFT--ILLTLWTMRCSTPSG 19
 ||| :||| : :|||
 Db 56 LFTGPLLLTLHHLLCETSPSG 76

RESULT 7

```

US-10-156-761-10847
; Sequence 10847, Application US/10156761
; Publication No. US20030119018A1
;
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: SHIBATA, HIROSHI
; APPLICANT: SHIBATA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
;
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
;
; SEQ ID NO 10847
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-10847

```

Query Match 46.5%; Score 47; DB 14; Length 252;
Best Local Similarity 57.1%; Pred. No. 67;
Matches 8; Conservative 2; Mismatches 4; Indels

Qy 6 LTLWTMRCSSTPSG 19
:|||||:||||
Db 235 VTLWTRRTAHTPGG 249

RESIST &

```

RES001 8
US-10-425-115-213984
; Sequence 213984, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222) B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 213984
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_12674C.1.pep
US-10-425-115-213984

```

Query Match 45.5%; Score 46; DB 17; Length 73;

```

Best Local Similarity  51.5%;  Pred. No. 29;
Matches 8;  Conservative 1;  Mismatches 4;  Indels 0;  Gaps 0;

Qy  5  LTLWTMRCSSTP 17
      | : |||||
Db  47  LTSRWTSRCSTP 59

```

RESULT 9

```

US-09-764-891-3368
; Sequence 3368, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3368
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring
; NAME/KEY: SITE
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals any of the naturally occurring
; NAME/KEY: SITE
; LOCATION: (70)
; OTHER INFORMATION: Xaa equals any of the naturally occurring
US-09-764-891-3368

```

```
Query Match      45.5%; Score 46; DB 10; Length 80;
Best Local Similarity 72.7%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

Qy 5 LLTLWTMRCSS 15
||| | : |||
Dp 38 LLCLWNMKCSS 48

RESIST. T 10

```

US-10-437-963-121796
; Sequence 121796, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 121796
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24786C.1.pep
US-10-437-963-121796

```

Query Match 45.5%; Score 46; DB 16; Length 98;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 ILLTLWTMRCSTPSG 19
... ||| ||| |||
Db 64 VVSWTMESSLPSC 79

RESULT 11
US-10-437-963-115917
; Sequence 115917, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 115917
; LENGTH: 1478
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: unsure
; LOCATION: (1)..(1478)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_19469C.1.pap
US-10-437-963-115917

Query Match 45.5%; Score 46; DB 16; Length 1478;
Best Local Similarity 53.8%; Pred. No. 5.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 TLWTMRCSTPSG 19
... ||| ||| |||
Db 220 TPWTARCGAAPAG 232

RESULT 12
US-10-424-599-181785
; Sequence 181785, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 181785
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_135165C.1.pap
US-10-424-599-181785

Query Match 44.6%; Score 45; DB 15; Length 55;

Best Local Similarity 43.8%; Pred. No. 31;
Matches 7; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 1 LFTILLTLWTMRCST 16
... ||| ||| |||
Db 38 LASVILAIPTVCTST 53

RESULT 13
US-10-437-963-127870
; Sequence 127870, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 127870
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_30279C.1.pap
US-10-437-963-127870

Query Match 44.6%; Score 45; DB 16; Length 84;
Best Local Similarity 53.3%; Pred. No. 46;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LFTILLTLWTMRCSS 15
... ||| ||| |||
Db 5 LIFITLLMYINCSS 19

RESULT 14
US-10-437-963-166688
; Sequence 166688, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 166688
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_65372C.1.pap
US-10-437-963-166688

Query Match 44.6%; Score 45; DB 16; Length 297;
Best Local Similarity 42.1%; Pred. No. 1.6e+02;

Search completed: November 10, 2004, 16:36:01
Job time : 19.3256 secs

This Page Blank (uspto)

Result No.	Score	Query Match	Length	DB	ID	Description
1	46	45.5	186	2	AD0357	probable exported
2	46	45.5	263	2	D82441	conserved hypochet
3	45.5	45.0	255	2	AH0577	molybdopterin-cont
4	45	44.6	260	2	T15432	hypothetical prote
5	45	44.6	560	2	E65083	Putative L-lactate
6	44	43.6	57	1	QXBp6L	hypothetical prote
7	44	43.6	119	2	T16950	hypothetical prote
8	44	43.6	432	2	A34413	atrial gland granu
9	43	42.6	82	2	T42312	hypothetical prote
10	43	42.6	148	2	T12798	hypothetical prote
11	43	42.6	290	2	S78540	hypothetical prote
12	43	42.6	301	2	E72566	hypothetical prote
13	43	42.6	493	2	B97724	hypothetical prote
14	43	42.6	493	2	E71725	glutamyl-tRNA amid
15	43	42.6	512	2	H84698	glutamyl-tRNA amid
16	42	41.6	221	2	JC7587	stromal cell-deriv
17	42	41.6	397	2	G90667	probable transport
18	42	41.6	397	2	C85518	probable transcript
19	42	41.6	409	2	H83973	hypothetical prote
20	42	41.6	721	2	AH3417	lpsa protein [impo
21	42	41.6	1685	2	T02750	acetyl-CoA carboxy
22	41	40.6	123	2	S29121	GTP-binding regula
23	41	40.6	182	2	C89481	protein R04B3.3 (i
24	41	40.6	206	1	A55169	pre-T-cell recepto
25	41	40.6	215	2	A10713	Probable membrane
26	41	40.6	326	1	RGSOB2	GTP-binding regula
27	41	40.6	340	1	RGBOB1	GTP-binding regula
28	41	40.6	340	1	RGFfBH	GTP-binding regula
29	41	40.6	340	1	RGHUB1	GTP-binding regula

Best Local Similarity 47.1%; Pred. No. 12;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 3 TLLTLTLMRCSSPTSG 19
DB 9 TLLTLTGLACASPTG 25

RESULT 3
AH0577
molybdopterin-containing oxidoreductase membrane anchor chain STY0661 [imported] - Salmonella enterica subsp. enterica serovar Typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0577
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Mouton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi
A:Reference number: AB0502; MUID:21534947; PMID:11577608
A:Accession: AH0577
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05090.1; PID:g16501865; GSPDB:GN00176
C:Genetics:
A:Gene: STY0661
C:Superfamily: probable dimethylsulfoxide reductase chain C

Query Match 45.0%; Score 45.5; DB 2; Length 255;
Best Local Similarity 47.4%; Pred. No. 14;
Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

QY 2 FTLL--TLTMRCSSTP 17
DB 134 FTMLFFITVWVGCAAP 152

RESULT 4
T15432
hypothetical protein C06E2.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15432
R:Favella, T.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid C06E2.
A:Reference number: Z18349
A:Accession: T15432
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-260 <FAV>
A:Cross-references: UNIPROT:P52484; EMBL:U39674; NID:G1049421; PID:G1049424; PIDN:AAA804
C:Genetics:
A:Gene: CESP:C06E2.3
A:Introns: 54/3; 129/2; 171/3
C:Superfamily: yeast ubiquitin-protein ligase UBC1

Query Match 44.6%; Score 45; DB 2; Length 260;
Best Local Similarity 55.6%; Pred. No. 17;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 LFTILLTLTMRCSSTPS 18
DB 135 LRTVLLSQAMLCSPEPS 152

RESULT 5
E65083
Putative L-lactate permease - Escherichia coli (strain K-12)
C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: E65083
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E65083
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-560 <BLAT>
A:Cross-references: UNIPROT:Q46939; GB:AE000380; GB:U00096; NID:g1789344; PIDN:AAC76011.1
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: L-lactate permease

Query Match 44.6%; Score 45; DB 2; Length 560;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 LFTILLTLTMR 12
DB 316 ILTVLVTITMK 327

RESULT 6
QXBP6L
hypothetical protein A-57 (nin region) - phage lambda
C:Species: phage lambda
C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 09-Jul-2004
C:Accession: B43011; G43016; A04392
R:Daniels, D.
submitted to the Nucleic Acid Sequence Database, September 1982
A:Reference number: A94614
A:Accession: B43011
A:Molecule type: DNA
A:Residues: 1-57 <DAN>
A:Cross-references: UNIPROT:P03767
R:Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B.
J. Mol. Biol. 162, 729-773, 1982
A:Title: Nucleotide sequence of bacteriophage lambda DNA.
A:Reference number: A92891; MUID:83189071; PMID:622115
A:Accession: G43016
A:Molecule type: DNA
A:Residues: 1-9, 'C', 11-57 <SAN>
A:Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:G215104; I
C:Genetics:
A:Map position: 86.49-86.84
C:Superfamily: phage lambda hypothetical 7K protein (nin region)

Query Match 43.6%; Score 44; DB 1; Length 57;
Best Local Similarity 54.5%; Pred. No. 6.5;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 9 WTMRCSSPTSG 19
DB 26 WLCRCERTPTG 36

RESULT 7
T16950
hypothetical protein T28D9.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16950
R:Fulton, L.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid T28D9.
A:Reference number: Z18614
A:Accession: T16950
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-119 <FUL>
A:Cross-references: UNIPROT:Q10026; EMBL:U28738; NID:g861262; PID:g861266; PIDN:AAA68311.

A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP.T28D9.9
A;Introns: 32/1

Query Match 43.6%; Score 44; DB 2; Length 119;
Best Local Similarity 70.0%; Pred. NO. 12;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 LWTMRCSSTP 17
|||:|||||
DB 4 LWTQRCSTP 13

RESULT 8
A34413
atrial gland granule-specific antigen precursor - California sea hare
C;Species: Aplysia californica (California sea hare)
C;Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 09-Jul-2004
C;Accession: A34413
R;Sossin, W.S.; Kreiner, T.; Barinaga, M.; Schilling, J.; Scheller, R.H.
J. Biol. Chem. 264, 16933-16940, 1989
A;Title: A dense core vesicle protein is restricted to the cortex of granules in the exo
A;Reference number: A34413; MUID:89380331; PMID:2777814
A;Accession: A34413
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-432 <SOS>
A;Cross-references: UNIPROT:P15287; GB:J05059; NID:G155710; PID:G155710
C;Superfamily: California sea hare atrial gland granule-specific antigen

Query Match 43.6%; Score 44; DB 2; Length 432;
Best Local Similarity 61.5%; Pred. NO. 37;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 LLLTWTMRCSSTP 17
|||:|||||
DB 371 LTTTWTMPCSTP 383

RESULT 9
T42312
hypothetical protein - phage SPPI
C;Species: phage SPPI
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T42312
R;Alonso, J.C.; Luder, G.; Stiege, A.C.; Chai, S.; Weise, F.; Trautner, T.A.
Gene 204, 201-212, 1997
A;Title: The complete nucleotide sequence and functional organization of Bacillus subtil
A;Reference number: Z22137; MUID:98094274; PMID:9434185
A;Accession: T42312
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-82 <ALO>
A;Cross-references: UNIPROT:O48472; EMBL:X97918; PIDN:CAA66519.1

Query Match 42.6%; Score 43; DB 2; Length 82;
Best Local Similarity 70.0%; Pred. NO. 13;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LFTILLTLWT 10
|||:|||||
DB 48 LFTIVTWT 57

RESULT 10
T12798
hypothetical protein yomK - Bacillus subtilis phage SPBc2
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C;Accession: T12798; C69911
R;Lazarevic, V.; Dueterhoeft, A.; Soldo, B.; Hilbert, H.; Mael, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997

A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 prophag
A;Reference number: Z17583
A;Accession: T12798
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-148 <LAZ>
A;Cross-references: UNIPROT:O64047; EMBL:AF020713; NID:G3025478; PID:G3025512; PIDN:AAAC1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Broutillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerk
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.P.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: C69911
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-148 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:G2634478; PIDN:CAB14051.1; PID:G2634553
A;Experimental source: strain 168
C;Genetics:
A;Gene: yomK
C;Superfamily: Bacillus subtilis phage SPBc2 hypothetical protein yomK

Query Match 42.6%; Score 43; DB 2; Length 148;
Best Local Similarity 58.3%; Pred. NO. 21;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 LFTILLTLWTMR 12
|||:|||||
DB 110 LITIIICLWTR 121

RESULT 11
S74540
hypothetical protein sl10696 - Synecocystis sp. (strain PCC 6803)
C;Species: Synecocystis sp.
C;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S74540
R;Kansko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.

A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S74540
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-230 <KAN>
A;Cross-references: UNIPROT:P72685; EMBL:D90899; GB:AB001339; NID:G1651650; PIDN:BAAL669;
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 42.6%; Score 43; DB 2; Length 290;
Best Local Similarity 41.2%; Pred. NO. 37;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 LFTILLTLWTMRCSSTP 17
|||:|||||
DB 63 iFLVLLIFWSHCYQTP 79

RESULT 12
E72566

A;Gene: gata; RP152
C;Superfamily: indoleacetamide hydrolase

Query Match 42.6%; Score 43; DB 2; Length 493;
Best Local Similarity 47.4%; Pred. No. 59;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 LFTILLTLWTMRCSSSTPSG 19
||||| : : ||| :
Db 432 LFTIPASLAGLPCASVPAG 450

RESULT 15
H84698

hypothetical protein At2g29650 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: H84698
R;Lin. X.; Xaul. S.; Rounslev, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: H84698
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: UNIPROT:082390; GB:AE002093; NID:g3582333; PIDN:AAC35230.1; GSPDB:GN
C;Genetics:
A;Gene: At2g29650
A;Map position: 2
C;Superfamily: hexuronate transporter

Query Match 42.6%; Score 43; DB 2; Length 512;
Best Local Similarity 47.1%; Pred. No. 61;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 LFTILLTLWTMRCSSSTP 17
| : | : | : | : | : | :
Db 269 LCTVWLTLLWLTKAESSP 285

Search completed: November 10, 2004, 12:29:21
Job time : 4.84105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01 ; Search time 20.7886 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-17

Perfect score: 101

Sequence: 1 LFTLLTLWTMRCSTSPSG 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	53.5	315	2 Q8VGJ6	Q8VGJ6 mus musculus
2	54	53.5	318	2 Q7TQ53	Q7TQ53 mus musculus
3	48.5	48.0	946	2 Q91Y12	Q91Y12 mus musculus
4	48	47.5	231	1 CDTA HELHP	Q91Y16 helicobacte
5	48	47.5	391	2 Q7W0R6	Q7W0R6 bordetella
6	48	47.5	408	2 Q7W2G6	Q7W2G6 bordetella
7	48	47.5	410	2 Q7WRD6	Q7WRD6 bordetella
8	47	46.5	252	2 Q82148	Q82148 streptomyce
9	47	46.5	373	2 Q8ECG6	Q8ECG6 shewanella
10	47	46.5	555	2 Q9K438	Q9K438 streptomyce
11	47	46.5	558	2 Q7XW55	Q7XW55 cryza sativ
12	47	46.5	719	2 Q72G36	Q72G36 desulfobact
13	47	46.5	719	2 AAS94509	AAS94509 desulfovi
14	46	45.5	80	2 Q6ZLF8	Q6ZLF8 cryza sativ
15	46	45.5	80	2 BAC33009	BAC33009 cryza sat
16	46	45.5	139	2 Q8VJF0	Q8VJF0 mycobacteri
17	46	45.5	186	2 Q8ZCF2	Q8ZCF2 versinia pe
18	46	45.5	199	2 Q8CLD9	Q8CLD9 versinia pe
19	46	45.5	199	2 AAS62946	AAS62946 versinia
20	46	45.5	263	2 Q9KX08	Q9KX08 vibrio chol
21	46	45.5	321	2 Q8IYW0	Q8IYW0 homo sapien
22	46	45.5	321	2 AAQ88538	AAQ88538 homo sapi
23	46	45.5	380	2 Q9H8E9	Q9H8E9 homo sapien
24	46	45.5	947	2 Q8G7D8	Q8G7D8 bifidobacte
25	45.5	45.0	255	2 Q8Z8J6	Q8Z8J6 salmonella
26	45.5	45.0	255	2 Q8Z818	Q8Z818 salmonella
27	45	44.6	260	1 UB21 CAEEL	UB2184 caenorhabdi
28	45	44.6	341	2 Q7P860	Q7P860 anopheles g
29	45	44.6	382	2 Q7PDC6	Q7PDC6 anopheles g
30	45	44.6	418	2 Q9LAF4	Q9LAF4 bacillus sp
31	45	44.6	431	2 Q7UI89	Q7UI89 rhodopirell

32 45 44.6 446 2 Q9VJ92 Q9VJ92 drosophila
33 45 44.6 465 2 Q83Q71 Q83Q71 shigella fl
34 45 44.6 528 2 Q7UBL1 Q7UBL1 shigella fl
35 45 44.6 540 2 Q7T6H7 Q7T6H7 sugarcane y
36 45 44.6 560 1 GLCA_ECOLI Q46839 escherichia
37 45 44.6 560 2 Q8FDN7 Q8FDN7 escherichia
38 45 44.6 560 2 CAE85240 CAE85240 escherich
39 45 44.6 564 2 Q8TLQ4 Q8TLQ4 methanosarc
40 45 44.6 1003 2 Q91ZU9 Q91ZU9 mus musculu
41 44 43.6 57 1 NIND_LAMED P03767 bacterioph
42 44 43.6 119 1 YXK9 CAEEL Q10026 caenorhabdi
43 44 43.6 293 2 Q7Z4D3 Q7Z4D3 homo sapien
44 44 43.6 338 2 Q6P025 Q6P025 brachydanio
45 44 43.6 338 2 AAH59436 AAH59436 brachydan

ALIGNMENTS

RESULT 1

Q8VGJ6 PRELIMINARY; PRT; 315 AA.
AC Q8VGJ6;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Olfactory receptor MOR266-2 (Olfactory receptor Olfr1417).
GN Name=Olfr1417;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21676863; PubMed=11802173;
RA Zhang X., Firestein S.;
RT "The Olfactory receptor gene superfamily of the mouse.";
RL Nat. Neurosci. 5:124-133(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21864068; PubMed=11875048;
RA Young J.M., Friedman C., Williams E.M., Ross J.A., Tonnes-Priddy L.,
RA Trask B.J.;
RT "Different evolutionary processes shaped the mouse and human olfactory
receptor gene families.";
RL Hum. Mol. Genet. 11:535-546(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Adams M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22974002; PubMed=14611657;
RA Young J.M., Shykind B.M., Lane R.P., Tonnes-Priddy L., Ross J.A.,
RA Walker M., Williams E.M., Trask B.J.;
RT "Odorant receptor expressed sequence tags demonstrate olfactory
expression of over 400 genes, extensive alternate splicing and unequal
expression levels.";
RL Genome Biol. 4:R71-R71(2003).
RN [5]
RP SEQUENCE FROM N.A.
RA Sanders K.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
DR EMBL; AY073152; AAL60815.1; -.
DR EMBL; AY318635; AAP71795.1; -.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0004984; F: olfactory receptor activity; IEA.
DR GO; GO:0004872; F: receptor activity; IEA.
DR GO; GO:0007186; F: G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR000376; GPCR_Rhodpsn.
DR InterPro; IPR000725; Olfact_receptor.

```

DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00245; OLFACTORYR.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR_F1_1; 1.
DR PROSITE; PS00262; G-PROTEIN RECEPTOR_F1_2; 1.
DR PROSITE; PS00262; G-PROTEIN RECEPTOR; Transmembrane.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 315 AA; 35583 MW; A9C2A3E88501A179 CRC64;

Query Match 53.5%; Score 54; DB 2; Length 315;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 9; Conservative 3; Mismatches 0; Indels 3; Gaps 0;

QY 3 TILLTLWTMRCSSTP 17
   :||:|||||
DB 44 SIVLTINRCLHTP 58
   :||:|||||

RESULT 2
ID Q7TQ33 PRELIMINARY; PRT; 318 AA.
AC Q7TQ33.
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Olfactory receptor Olfr1418.
GN Name=Olfr1418;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22974002; PubMed=14611657;
RA Young J.M., Shykind B.M., Lane R.P., Tonnes-Priddy L., Ross J.A.,
RA Walker M., Williams E.M., Trask B.J.;
RT "Odorant receptor expressed sequence tags demonstrate olfactory
RT expression of over 400 genes, extensive alternate splicing and unequal
RT expression levels.";
RL Genome Biol. 4:R71-R71(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Sanders K.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SURCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
DR EMBL; AY318636; AAP71796.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin...; IEA.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR_F1_1; 1.
DR PROSITE; PS00262; G-PROTEIN RECEPTOR_F1_2; 1.
DR PROSITE; PS00262; G-PROTEIN RECEPTOR; Transmembrane.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 318 AA; 35737 MW; 548C4CA4EF8D20EB CRC64;

Query Match 53.5%; Score 54; DB 2; Length 318;
Best Local Similarity 60.0%; Pred. No. 4.7;
Matches 9; Conservative 3; Mismatches 0; Indels 3; Gaps 0;

QY 3 TILLTLWTMRCSSTP 17
   :||:|||||
DB 44 SIVLTINRCLHTP 58
   :||:|||||

RESULT 3
ID Q91V12 PRELIMINARY; PRT; 946 AA.
AC Q91V12.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

```

```

DE Protocadherin alpha 8.
GN Name=Pcdhal;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes.";
RL Cell 97:779-790(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202599; PubMed=10716726;
RA Wu Q., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic
RT feature of protocadherin genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21154914; PubMed=11230163;
RA Wu Q., Zhang T., Cheng J.-F., Kim Y., Grimwood J., Schmutz J.,
RA Dickson M., Noonan J.P., Zhang M.Q., Myers R.M., Maniatis T.;
RT "Comparative DNA sequence analysis of mouse and human protocadherin
RT gene clusters.";
RL Genome Res. 11:389-404(2001).
CC -!- SIMILARITY: Contains 6 cadherin domains.
DR EMBL; AY013766; AAK26055.1; -.
DR MGI; MGI:2150982; Pcdhal.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; Cadherin; 5.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 6.
DR PROSITE; PS00232; CADHERIN_1; 5.
DR PROSITE; PS00268; CADHERIN_2; 6.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Calcium; calcium-binding.
SQ SEQUENCE 946 AA; 101759 MW; 5E060D4A81E1076F CRC64;

Query Match 48.0%; Score 48.5; DB 2; Length 946;
Best Local Similarity 50.0%; Pred. No. 93;
Matches 10; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY 1 LFTLLTLWT-MRCSSTPSG 19
   :||:|||||
DB 710 LVLTLTLTLALRCSATPTG 729
   :||:|||||

RESULT 4
ID CDTA_HELP STANDARD; PRT; 231 AA.
AC Q9RFY6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Cytolethal distending toxin subunit A precursor (CDT A).
GN Name=cdtA; OrderedLocusNames=HH1446;
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;

```

```

RX MEDLINE=20072683; PubMed=10603386;
RA Young V.B., Knox K.A., Schauer D.B.;
RT "Cytolethal distending toxin sequence and activity in the
RL enterohepatic pathogen Helicobacter hepaticus.";
RN Infect. Immun. 68:184-191(2000).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954; DOI=10.1073/pnas.1332093100;
RA Suerbaum S., Josephans C., Stenzenbach T., Drescher B., Brandt P.,
RA Ball M., Droge M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Venz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
CC -!- FUNCTION: CDTs are cytotoxins which induce cell distension, growth
CC arrest in G2/M phase, nucleus swelling, and chromatin
CC fragmentation in HeLa cells.
CC -!- SUBUNIT: Composed of three subunits; A, B and C.
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF163667; AAP19157.1; -.
DR EMBL; AE017148; AAP78043.1; -.
DR InterPro; IPR003558; CDTxinA.
DR InterPro; IPR008997; RicinB like.
DR InterPro; IPR000772; Ricin B lectin.
DR Pfam; PF03498; CDTxinA; 1.
DR PRINTS; PR01387; CDTXINA.
DR PROSITE; PS00033; PROXAR_LIPOPROTEIN; 1.
DR PROSITE; PS0231; RICIN B LECTIN; 1.
DR Complete proteome; Lectin; Lipoprotein; Outer membrane; Palmitate;
KW Signal; Toxin.
FT SIGNAL 1 15 Potential.
FT CHAIN 16 231 Cytolethal distending toxin subunit A.
FT DOMAIN 130 217 Ricin B-type lectin.
FT LIPID 16 16 N-palmitoyl cysteine (Potential).
FT LIPID 16 16 S-diacylglycerol cysteine (Potential).
SQ SEQUENCE 231 AA; 26021 MW; 55E3ED5FC6E54D86 CRC64;

Query Match 47.5%; Score 48; DB 1; Length 231;
Best Local Similarity 59.8%; Pred. No. 29;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 LFTILLTWTWRCSTP 17
Db 4 LFFLLTLLFAACSTP 20

RESULT 5
Q7W0R6 PRELIMINARY; PRT; 391 AA.
AC Q7W0R6;
DT 01-OCT-2003 (TrEMBLrel. 25; Created)
DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
DE Putative membrane protein.
GN OrderedLocNames-BP0018;
OS Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]

```

```

SEQUENCE FROM N.A.
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jegels K.,
RA Leather S., Mouton S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640411; CAE39758.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot_kinase.
DR ProDom; PD000001; Prot_kinase; 1.
KW Complete proteome.
SQ SEQUENCE 391 AA; 40841 MW; 31AAEA8BF1F4B69 CRC64;

Query Match 47.5%; Score 48; DB 2; Length 391;
Best Local Similarity 44.4%; Pred. No. 48;
Matches 12; Conservative 1; Mismatches 6; Indels 8; Gaps 1;

Qy 1 LFTILLT-----WTMRCSTSPG 19
Db 224 LFAILLALATLGVAVYMWQRLTGPSTG 250

RESULT 6
Q7W2G6 PRELIMINARY; PRT; 408 AA.
AC Q7W2G6;
DT 01-OCT-2003 (TrEMBLrel. 25; Created)
DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
DE Putative membrane protein.
GN OrderedLocNames-BP0017;
OS Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jegels K.,
RA Leather S., Mouton S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640423; CAE39758.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot_kinase.
DR ProDom; PD000001; Prot_kinase; 1.
KW Complete proteome.
SQ SEQUENCE 408 AA; 42707 MW; A84738FA94AE59DD CRC64;

```

Query Match 47.5%; Score 48; DB 2; Length 408;
 Best Local Similarity 44.4%; Pred. No. 50;
 Matches 12; Conservative 1; Mismatches 6; Indels 8; Gaps 1;

QY 1 LFTILLTL-----WTMRCSTSPSG 19
 ||||| : : : : :
 Db 228 LFAILLALALGVAVYVWQRLTGTPSG 254

RESULT 7

QYWRD6 PRELIMINARY; PRT; 410 AA.
 AC Q7WRD6;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative membrane protein.
 GN OrderedLocusNames=B80017;
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagers K.,
 RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,
 RT Bordetella parapertussis and Bordetella bronchiseptica.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640437; CAE30519.1; -;
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004672; F:protein kinase activity; IEA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
 DR InterPro: IPR000719; Prot. kinase.
 DR ProDom: PD000001; Prot_kinase; 1.
 KW Complete proteome.
 SQ SEQUENCE 410 AA; 42806 MW; 0BDC294518BD02A4 CRC64;

Query Match 47.5%; Score 48; DB 2; Length 410;
 Best Local Similarity 44.4%; Pred. No. 50;
 Matches 12; Conservative 1; Mismatches 6; Indels 8; Gaps 1;

QY 1 LFTILLTL-----WTMRCSTSPSG 19
 ||||| : : : : :
 Db 228 LFAILLALALGVAVYVWQRLTGTPSG 254

RESULT 8

ID Q82148 PRELIMINARY; PRT; 252 AA.
 AC Q82148;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocusNames=SAV310;
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680;

RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Morikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005034; BAC71021.1; -;
 DR InterPro: IPR011046; WD40_like.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 252 AA; 26338 MW; 87AAF9911EB6F085 CRC64;

Query Match 46.5%; Score 47; DB 2; Length 252;
 Best Local Similarity 57.1%; Pred. No. 46;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 LFTLWTMRCSTSPSG 19
 :|||: :|||:
 Db 235 VTLWTRTAHTPGG 248

RESULT 9

QYECG6 PRELIMINARY; PRT; 373 AA.
 ID Q8EGC6;
 AC Q8EGC6;
 DT 01-VAR-2003 (TREMBlrel. 23, Created)
 DT 01-VAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
 DE Glycosyl transferase, group 1 family protein.
 GN OrderedLocusNames=SQ3176;
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813; DOI=10.1038/nbt749;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N.L., Methe B.A.,
 RA Clayton R.A., Meyer T., Teapin A., Scott J., Beanan M.J.,
 RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
 RA Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,
 RA White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Impraim M.,
 RA Lee K., Berry K.J., Lee C., Mueller J., Khouri H.M., Gill J.,
 RA Utterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O.,
 RA Venter J.C., Nealon K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AE015755; AAN56176.1; -;
 DR TIGR; SO3176; -;
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR InterPro: IPR01296; Glyco_transf_1.
 DR Pfam; PF00534; Glycos_transf_1; 1.
 KW Complete proteome; Transferase.
 SQ SEQUENCE 373 AA; 41298 MW; 51865EF2E6C4B95A CRC64;

Query Match 46.5%; Score 47; DB 2; Length 373;
 Best Local Similarity 50.0%; Pred. No. 66;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

QY      8 LWTMRCSSTPSG 19
Db      108 IWSVRCNSNPAG 119

RESULT 10
Q9K438 ID Q9K438 PRELIMINARY; PRT; 555 AA.
AC Q9K438;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative bifunctional protein (secreted sugar binding protein/sugar
DE hydrolase).
GN OrderedLocusNames=SC01061; ORFNames=SCG22.07, SCG22.07c;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21956410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR ENBL; AL339107; CAB95280.1; -.
DR HSP. P26514; IKNM.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR InterPro; IPR002860; Glyco_hydro_BMR.
DR InterPro; IPR008997; RicinB_like.
DR InterPro; IPR000772; Ricin B lectin.
DR InterPro; IPR011040; Sialidase.
DR Pfam; PF02012; BNR; 3.
DR Pfam; PF00652; Ricin B lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LLECTIN; 1.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 555 AA; 58835 MW; E78C9A88990B2786 CRC64;

Query Match 46.5%; Score 47; DB 2; Length 555;
Best Local Similarity 47.4%; Pred. No. 95;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY      1 LFTLLTWTMRCSSTPSG 19
Db      20 LLTALLSLWSQASAPVG 38

RESULT 11
Q7XNW5 ID Q7XNW5 PRELIMINARY; PRT; 558 AA.
AC Q7XNW5;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE OSUNB50040D15.9 protein.
GN Name=OSUNB50040D15.9;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX PubMed=12447439;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
DR ENBL; AL60655; CAB04419.2; -.
DR Gramene; Q7XNW5; -.
SQ SEQUENCE 558 AA; 61560 MW; CB9ADAE249570AC8 CRC64;

Query Match 46.5%; Score 47; DB 2; Length 558;
Best Local Similarity 35.3%; Pred. No. 96;
Matches 6; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY      3 TILLTWTMRCSSTPSG 19
Db      90 TLLCTVWALKCANAG 106

RESULT 12
Q7G36 ID Q7G36 PRELIMINARY; PRT; 719 AA.
AC Q7G36;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Sensory box histidine kinase.
GN OrderedLocusNames=DVU00025;
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB
OS 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15077118; DOI=10.1038/nbt959;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N.I., Methe B.A., Brinkac L.M.,
RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidse I.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Knouri H.M., Gill J., Utterback T.R.,
RA Felldlyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT Desulfovibrio vulgaris Hildenborough.";
RL Nat. Biotechnol. 22:554-559(2004).
CC -!- SIMILARITY: Contains 1 histidine kinase domain.
DR ENBL; AB017309; AAS94509.1; -.
DR TIGR; DVU0025; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR004358; Bact_sens_pr_C.
DR InterPro; IPR005467; His_Kinase.
DR InterPro; IPR003661; His_kin_N.
DR InterPro; IPR009082; His_kin_Homodim.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS.
DR InterPro; IPR000700; PAS-assoc_C.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00512; HSKA; 1.
DR Pfam; PF00785; PAC; 2.
DR PRINTS; PR00344; BCTRLSENSOR.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00388; HSKA; 1.
DR SMART; SM00086; PAC; 2.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 8.33951 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-18

Perfect score: 136
Sequence: 1 AGEDMEISVKELRTILNRIISKHDLRT 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 56318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pdp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pdp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pdp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pdp.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pdp.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	136	100.0	714	3	US-09-422-869-22
2	136	100.0	714	4	US-09-538-092-867
3	128	94.1	714	4	US-09-308-345A-47
4	95	69.9	757	4	US-09-622-880B-1
5	95	69.9	821	4	US-09-308-345A-48
6	95	69.9	821	4	US-09-622-880B-15
7	94	69.1	821	3	US-09-422-869-24
8	94	69.1	821	4	US-09-538-092-972
9	92	67.6	248	3	US-09-032-523-1
10	92	67.6	248	4	US-09-802-633-1
11	91	66.9	266	3	US-09-032-523-7
12	91	66.9	266	4	US-09-802-633-7
13	91	66.9	268	2	US-08-835-099A-9
14	91	66.9	268	3	US-09-157-349-9
15	81	59.6	251	1	US-08-726-525-6
16	81	59.6	251	2	US-08-487-942-6
17	81	59.6	251	3	US-08-726-036A-6
18	81	59.6	251	3	US-09-083-516-6
19	81	59.6	700	1	US-08-726-525-7
20	81	59.6	700	2	US-08-487-942-7
21	81	59.6	700	2	US-08-726-036A-7
22	81	59.6	700	3	US-09-422-869-23
23	81	59.6	700	3	US-09-083-516-7
24	81	59.6	700	4	US-09-308-345A-46
25	81	59.6	700	4	US-09-538-092-947
26	79	58.1	664	4	US-09-377-497-70
27	79	58.1	690	3	US-09-422-869-28

28	68	50.0	703	3	US-09-422-869-27	Sequence 27, Appli
29	64	47.1	686	4	US-09-653-839-8	Sequence 8, Appli
30	64	47.1	686	4	US-10-202-619-8	Sequence 8, Appli
31	64	47.1	702	4	US-09-653-839-6	Sequence 6, Appli
32	64	47.1	702	4	US-10-202-619-6	Sequence 6, Appli
33	64	47.1	703	4	US-09-308-345A-49	Sequence 49, Appli
34	64	47.1	723	4	US-09-653-839-4	Sequence 4, Appli
35	64	47.1	723	4	US-10-202-619-4	Sequence 4, Appli
36	64	47.1	739	4	US-09-653-839-2	Sequence 2, Appli
37	64	47.1	739	4	US-10-202-619-2	Sequence 1, Appli
38	59	43.4	703	2	US-08-835-099A-1	Sequence 1, Appli
39	59	43.4	703	3	US-09-157-349-1	Sequence 2, Appli
40	59	43.4	712	3	US-08-835-099A-2	Sequence 2, Appli
41	59	43.4	712	3	US-09-157-349-2	Sequence 2, Appli
42	54	39.7	716	4	US-09-248-796A-24140	Sequence 24140, A
43	52	38.2	675	6	5258288-1	Patent No. 5258288
44	51	37.5	34	3	US-08-978-285-2	Sequence 2, Appli
45	51	37.5	34	3	US-08-978-285-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-422-869-22
; Sequence 22, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG I.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-22

Query Match 100.0%; Score 136; DB 3; Length 714;
Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHDLRT 28
DB 555 AGEDMEISVKELRTILNRIISKHDLRT 582

RESULT 2
US-09-538-092-867
; Sequence 867, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01

```
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 867
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P07384
US-09-538-092-867

Query Match      100.0%; Score 136; DB 4; Length 714;
Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db 555 AGEDMEISVKELRTILNRIISKHKDLRT 582

RESULT 3
US-09-308-345A-47
; Sequence 47, Application US/09308345A
; Patent No. 6569665
; GENERAL INFORMATION:
; APPLICANT: BOEHM, Thomas;
; APPLICANT: DEAR, Neil T.
; TITLE OF INVENTION: No. 6569665el calpains, their preparation and use
; FILE REFERENCE: 0050/47576
; CURRENT APPLICATION NUMBER: US/09/308,345A
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: WordPerfect v. 6.1
; SEQ ID NO 47
; LENGTH: 714
; TYPE: PRT
; ORGANISM: human
US-09-308-345A-47

Query Match      94.1%; Score 128; DB 4; Length 714;
Best Local Similarity 96.4%; Pred. No. 6.1e-11;
Matches 27; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db 555 AGEDMEISVKELRTILNRIISKHKDLRT 582

RESULT 4
US-09-622-880B-1
; Sequence 1, Application US/09622880B
; Patent No. 6582932
; GENERAL INFORMATION:
; APPLICANT: FUKIAGE, Chiho
; APPLICANT: AZUMA, Mitsuoyoshi
; TITLE OF INVENTION: NOVEL CALPIN AND DNA ENCODING THE SAME
; FILE REFERENCE: 2000-1125A/MWC/00177
; CURRENT APPLICATION NUMBER: US/09/622,880B
; CURRENT FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: JP 10-049430
; PRIOR FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 1
; LENGTH: 757
; TYPE: PRT
; ORGANISM: Sprague-Dawley rat
US-09-622-880B-1

Query Match      69.9%; Score 95; DB 4; Length 757;
Best Local Similarity 57.1%; Pred. No. 6.2e-06;
Matches 16; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db 662 AGDDMEICADELKNVLTNTVVKHKDLKT 689

RESULT 5
US-09-308-345A-48
; Sequence 48, Application US/09308345A
; Patent No. 6569665
; GENERAL INFORMATION:
; APPLICANT: BOEHM, Thomas;
; APPLICANT: DEAR, Neil T.
; TITLE OF INVENTION: No. 6569665el calpains, their preparation and use
; FILE REFERENCE: 0050/47576
; CURRENT APPLICATION NUMBER: US/09/308,345A
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: WordPerfect v. 6.1
; SEQ ID NO 48
; LENGTH: 821
; TYPE: PRT
; ORGANISM: mouse
US-09-308-345A-48

Query Match      69.9%; Score 95; DB 4; Length 821;
Best Local Similarity 57.1%; Pred. No. 6.8e-06;
Matches 16; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db 662 AGDDMEICADELKNVLTNTVVKHKDLKT 689

RESULT 6
US-09-622-880B-15
; Sequence 15, Application US/09622880B
; Patent No. 6582932
; GENERAL INFORMATION:
; APPLICANT: FUKIAGE, Chiho
; APPLICANT: AZUMA, Mitsuoyoshi
; TITLE OF INVENTION: NOVEL CALPIN AND DNA ENCODING THE SAME
; FILE REFERENCE: 2000-1125A/MWC/00177
; CURRENT APPLICATION NUMBER: US/09/622,880B
; CURRENT FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: JP 10-049430
; PRIOR FILING DATE: 1998-03-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 15
; LENGTH: 821
; TYPE: PRT
; ORGANISM: No. 6582932way rat
; FEATURE:
; OTHER INFORMATION: p94 protein
US-09-622-880B-15

Query Match      69.9%; Score 95; DB 4; Length 821;
Best Local Similarity 57.1%; Pred. No. 6.8e-06;
Matches 16; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db 662 AGDDMEICADELKNVLTNTVVKHKDLKT 689

RESULT 7
US-09-422-869-24
; Sequence 24, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
```



```

/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA: US/09/802,633
/ FILING DATE: 08-Mar-2001
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/032,523
/ FILING DATE: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0479 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-845-4166
/ TELEX: <Unknown>
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 248 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: KRRANOT01
/ CLONE: 456855
/ SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-802-633-1

```

```

Query Match 67.6%; Score 92; DB 4; Length 248;
Best Local Similarity 57.1%; Pred. No. 4.7e-06;
Matches 16; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 AGDMEISVKELRTILNRIISKHKDLRT 28
DB 89 AGPDMEVGATDLNINLVSKHKDLKT 116

```

```

RESULT 11
US-09-032-523-7
/ Sequence 7, Application US/09032523
/ Patent No. 6232454
/ GENERAL INFORMATION:
/ APPLICANT: Bandman, Olga
/ APPLICANT: Hillman, Jennifer L.
/ APPLICANT: Corley, Neil C.
/ APPLICANT: Guegler, Karl
/ APPLICANT: Baugh, Mariah
/ TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.
/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/032,523
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J
/ REGISTRATION NUMBER: 36,749

```

```

/ REFERENCE/DOCKET NUMBER: PF-0479 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-855-0555
/ TELEX: 650-845-4166
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 266 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 164403
US-09-032-523-7

```

```

Query Match 66.9%; Score 91; DB 3; Length 266;
Best Local Similarity 53.6%; Pred. No. 7.2e-06;
Matches 15; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 1 AGDMEISVKELRTILNRIISKHKDLRT 28
DB 107 AGDDMEVSATLWNLNKVTRPDLKT 134

```

```

RESULT 12
US-09-802-633-7
/ Sequence 7, Application US/09802633
/ Patent No. 6627605
/ GENERAL INFORMATION:
/ APPLICANT: Bandman, Olga
/ APPLICANT: Hillman, Jennifer L.
/ APPLICANT: Corley, Neil C.
/ APPLICANT: Guegler, Karl
/ APPLICANT: Baugh, Mariah
/ TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.
/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/802,633
/ FILING DATE: 08-Mar-2001
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/032,523
/ FILING DATE: <Unknown>
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0479 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-855-0555
/ TELEX: 650-845-4166
/ TELEX: <Unknown>
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 266 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GenBank
/ CLONE: 164403

```

```
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-802-633-7
    Query Match          66.9%; Score 91; DB 4; Length 266;
    Best Local Similarity 53.6%; Pred. No. 7.2e-06;
    Matches 15; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY      1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db      107 AGDDMEVSATELMNLKVNKVTTRHPDLKT 134
      |||:|:| |||:|:| |||:|:| |||:|:|
      |||:|:| |||:|:| |||:|:| |||:|:|

RESULT 13
US-08-835-099A-9
; Sequence 9, Application US/08835099A
; Patent No. 5874277
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasuishi
; APPLICANT: NISHI, Kazunori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,099A
; FILING DATE: 04-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 083649/1996
; FILING DATE: 05-APR-1996
; APPLICATION NUMBER: 97105508.2
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 268 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-835-099A-9

Query Match          66.9%; Score 91; DB 2; Length 268;
Best Local Similarity 53.6%; Pred. No. 7.2e-06;
Matches 15; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY      1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db      109 AGDDMEVSATELMNLKVNKVTTRHPDLKT 136
      |||:|:| |||:|:| |||:|:| |||:|:|
      |||:|:| |||:|:| |||:|:| |||:|:|

RESULT 14
US-09-157-349-9
; Sequence 9, Application US/09157349
```

```
; Patent No. 6068990
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasuishi
; APPLICANT: NISHI, Kazunori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/157,349
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/835,099
; FILING DATE:
; APPLICATION NUMBER: 97105508.2
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 268 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-157-349-9

Query Match          66.9%; Score 91; DB 3; Length 268;
Best Local Similarity 53.6%; Pred. No. 7.2e-06;
Matches 15; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY      1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db      109 AGDDMEVSATELMNLKVNKVTTRHPDLKT 136
      |||:|:| |||:|:| |||:|:| |||:|:|
      |||:|:| |||:|:| |||:|:| |||:|:|

RESULT 15
US-08-726-525-6
; Sequence 6, Application US/08726525
; Patent No. 5789181
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; TITLE OF INVENTION: BINDING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,525
FILING DATE: 07-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,942
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 251 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-726-525-6

Query Match 59.6%; Score 81; DB 1; Length 251;
Best Local Similarity 53.6%; Pred. No. 0.00022;
Matches 15; Conservative 8; Mismatches 5; Indels 0; Gaps 0;
QY 1 AGEDMEISVKELRTILNRIISKHDLRT 28
Db 93 AGEDAEISAFELQTLIRVLAKRQDIKS 120

Search completed: November 10, 2004, 12:32:30
Job time : 8.33951 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 10, 2004, 15:53:52 ; Search time 27.0062 Seconds
(without alignments)
366.225 Million cell updates/sec

Title: US-10-092-750-18
Perfect score: 136
Sequence: 1 AGEDMEISVKELRTILNRIISKHKDLRT 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	28	14	US-10-092-750-18
2	136	100.0	714	9	US-09-768-877-22
3	136	100.0	714	14	US-10-116-519-6
4	136	100.0	714	15	US-10-390-585-3
5	136	100.0	714	16	US-10-408-765A-1337
6	129	94.9	713	14	US-10-116-519-8
7	129	94.9	713	15	US-10-390-585-11
8	95	69.3	709	14	US-10-116-519-9
9	95	69.9	709	15	US-10-051-874-117
10	95	69.9	821	15	US-10-080-334-260
11	94	69.1	729	15	US-10-080-334-259
12	94	69.1	767	14	US-10-274-639-1
13	94	69.1	767	15	US-10-333-574-1

14	69.1	773	15	US-10-080-334-86	Sequence 86, Appl
15	69.1	773	15	US-10-336-472-122	Sequence 122, App
16	69.1	815	15	US-10-080-334-257	Sequence 257, App
17	69.1	815	15	US-10-080-334-258	Sequence 258, App
18	69.1	821	9	US-09-768-877-24	Sequence 24, Appl
19	69.1	821	14	US-10-116-519-7	Sequence 7, Appli
20	69.1	821	15	US-10-080-334-256	Sequence 256, App
21	69.1	821	15	US-10-390-585-5	Sequence 5, Appli
22	69.1	821	15	US-10-336-472-124	Sequence 124, App
23	67.6	248	15	US-10-621-263-1	Sequence 1, Appli
24	66.9	207	9	US-09-925-301-1001	Sequence 1001, Ap
25	66.9	247	15	US-10-275-505-4	Sequence 4, Appli
26	66.9	266	15	US-10-621-263-7	Sequence 7, Appli
27	66.9	442	15	US-10-467-042-14	Sequence 14, Appl
28	59.6	206	14	US-10-043-487-221	Sequence 221, App
29	59.6	251	9	US-09-840-707A-8	Sequence 8, Appli
30	59.6	251	10	US-09-884-319-6	Sequence 6, Appli
31	59.6	251	14	US-10-038-557A-8	Sequence 8, Appli
32	59.6	283	15	US-10-108-260A-4215	Sequence 4215, Ap
33	59.6	699	15	US-10-390-585-19	Sequence 19, Appl
34	59.6	700	9	US-09-840-707A-9	Sequence 9, Appli
35	59.6	700	9	US-09-768-877-23	Sequence 23, Appl
36	59.6	700	10	US-09-884-319-7	Sequence 7, Appli
37	59.6	700	14	US-10-038-557A-9	Sequence 9, Appli
38	59.6	700	14	US-10-116-519-12	Sequence 12, Appl
39	59.6	700	15	US-10-390-585-4	Sequence 4, Appli
40	59.6	700	15	US-10-390-585-12	Sequence 12, Appl
41	59.6	700	16	US-10-408-765A-390	Sequence 390, App
42	58.8	702	15	US-10-051-874-114	Sequence 114, App
43	59.6	664	14	US-10-325-881-70	Sequence 70, Appl
44	58.1	690	9	US-09-768-877-28	Sequence 28, Appl
45	58.1	690	14	US-10-116-519-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-10-092-750-18
; Sequence 18, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-XI
; FILE REFERENCE: 500367050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; TYPE: PRT
; LENGTH: 28
; ORGANISM: Homo sapiens
US-10-092-750-18

Query Match 100.0%; Score 136; DB 14; Length 28;
Best Local Similarity 100.0%; Pred. No. 1.5e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AGEDMEISVKELRTILNRIISKHKDLRT	28
Db	1	AGEDMEISVKELRTILNRIISKHKDLRT	28

RESULT 2
US-09-768-877-22
; Sequence 22, Application US/09768877
; Patent No. US20020150896A1
; GENERAL INFORMATION:


```
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/296,180
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/300,620
; PRIOR FILING DATE: 2001-06-25
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-116-519-8

Query Match          94.9%; Score 129; DB 14; Length 713;
Best Local Similarity 92.9%; Pred. No. 7.3e-10;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGDMEISVKELRTILNRIISKHKDLRT 28
Db      554 AGDDMEISVKELQTLNRIISKHKDLRT 581

RESULT 7
US-10-390-585-11
; Sequence 11, Application US/10390585
; Publication No. US20040014093A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPAIN
; FILE REFERENCE: D0219ND
; CURRENT APPLICATION NUMBER: US/10/390,585
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: U.S. 60/364,941
; PRIOR FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-390-585-11

Query Match          94.9%; Score 129; DB 15; Length 713;
Best Local Similarity 92.9%; Pred. No. 7.3e-10;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGDMEISVKELRTILNRIISKHKDLRT 28
Db      554 AGDDMEISVKELQTLNRIISKHKDLRT 581

RESULT 8
US-10-116-519-9
; Sequence 9, Application US/10116519
; Publication No. US20030114373A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPAIN
; FILE REFERENCE: D0124 NP
; CURRENT APPLICATION NUMBER: US/10/116,519
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: US 60/281,253
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/288,768
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/296,180
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/300,620
; PRIOR FILING DATE: 2001-06-25
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 9
; LENGTH: 709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-519-9

Query Match          69.9%; Score 95; DB 14; Length 709;
Best Local Similarity 57.1%; Pred. No. 5.9e-05;
Matches 16; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY      1 AGDMEISVKELRTILNRIISKHKDLRT 28
Db      550 AGDDMEICADELKNVNTVVRKHKDLKT 577

RESULT 9
US-10-051-874-117
; Sequence 117, Application US/10051874
; Publication No. US20040005557A1
; GENERAL INFORMATION:
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shinkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glennda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
```


PRIOR FILLING DATE: 2000-07-28
 PRIOR APPLICATION NUMBER: US 60/223,544
 PRIOR FILLING DATE: 2000-08-04
 PRIOR APPLICATION NUMBER: US 60/224,717
 PRIOR FILLING DATE: 2000-08-11
 PRIOR APPLICATION NUMBER: US 60/225,988
 PRIOR FILLING DATE: 2000-08-16
 PRIOR APPLICATION NUMBER: US 60/227,568
 PRIOR FILLING DATE: 2000-08-23
 NUMBER OF SEQ ID NOS: 42
 SOFTWARE: PERL Program
 SEQ ID NO 1
 LENGTH: 767
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc feature
 OTHER INFORMATION: Incyte ID No. US20030232349A1 5155802CD1
 US-10-274-639-1

Query Match 69.1%; Score 94; DB 14; Length 767;
 Best Local Similarity 57.1%; Pred. No. 9,1e-05;
 Matches 16; Conservative 7; Mismatches 5; Indels

Qy 1 AGDMEISVKELRTILNRIISKHOLRT 28
 ||||| ||: ||: ||: ||: ||: ||:
 Db 608 AGDMEICADELKKVLNTVNVKHKOLKT 635

RESULT 13
 US-10-333-574-1
 Sequence 1, Application US/10333574
 Publication No. US20040091962A1
 GENERAL INFORMATION:
 APPLICANT: INCYTE GENOMICS, INC.
 APPLICANT: DELEGANE, Angelo M.; GANDHI, Ameena R.
 APPLICANT: HAFALIA, April J.A.; LU, Dyung Aina M.
 APPLICANT: ARVIZU, Chandra S.; TRIBOULEY, Catherine M.
 APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
 APPLICANT: NGUYEN, Danniell B.; LEE, Ernestine A.
 APPLICANT: KHAN, Farrah A.; YUE, Henry
 APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
 APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
 APPLICANT: YANG, Junming; THANGAVELU, Kavitha
 APPLICANT: DING, Li; KEARNEY, Liam
 APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.
 APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
 APPLICANT: BURFORD, Neil; CHAWLA, Narinder K.
 APPLICANT: LAL, Preeti G.; LEE, Sally
 APPLICANT: TODD, Stephen; LO, Terence P.
 APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
 APPLICANT: AZIMZAI, Yalda; LU, Yan
 TITLE OF INVENTION: PROTEASES
 FILE REFERENCE: PI-0167 USN
 CURRENT APPLICATION NUMBER: US/10/333,574
 CURRENT FILING DATE: 2003-01-21
 PRIOR APPLICATION NUMBER: US 01/22397
 PRIOR FILLING DATE: 2001-07-17
 PRIOR APPLICATION NUMBER: US 60/220,063
 PRIOR FILLING DATE: 2000-07-21
 PRIOR APPLICATION NUMBER: US 60/221,680
 PRIOR FILLING DATE: 2000-07-28
 PRIOR APPLICATION NUMBER: US 60/223,544
 PRIOR FILLING DATE: 2000-08-04
 PRIOR APPLICATION NUMBER: US 60/224,717
 PRIOR FILLING DATE: 2000-08-11
 PRIOR APPLICATION NUMBER: US 60/225,988
 PRIOR FILLING DATE: 2000-08-16
 PRIOR APPLICATION NUMBER: US 60/227,568
 NUMBER OF SEQ ID NOS: 42
 SOFTWARE: PERL Program
 SEQ ID NO 1

PRIOR APPLICATION NUMBER: 60/278,796
 PRIOR FILLING DATE: 2001-03-26
 PRIOR APPLICATION NUMBER: 60/281,521
 PRIOR FILLING DATE: 2001-04-04
 PRIOR APPLICATION NUMBER: 60/276,677
 PRIOR FILLING DATE: 2001-03-16
 PRIOR APPLICATION NUMBER: 60/311,595
 PRIOR FILLING DATE: 2001-08-10
 PRIOR APPLICATION NUMBER: 60/270,220
 PRIOR FILLING DATE: 2001-02-21
 PRIOR APPLICATION NUMBER: 60/274,295
 PRIOR FILLING DATE: 2001-03-08
 PRIOR APPLICATION NUMBER: 60/318,526
 PRIOR FILLING DATE: 2001-09-10
 PRIOR APPLICATION NUMBER: 60/286,548
 PRIOR FILLING DATE: 2001-04-25
 PRIOR APPLICATION NUMBER: 60/291,765
 PRIOR FILLING DATE: 2001-05-17
 PRIOR APPLICATION NUMBER: 60/270,797
 PRIOR FILLING DATE: 2001-02-23
 PRIOR APPLICATION NUMBER: 60/276,400
 PRIOR FILLING DATE: 2001-03-16
 PRIOR APPLICATION NUMBER: 60/270,810
 PRIOR FILLING DATE: 2001-02-23
 NUMBER OF SEQ ID NOS: 388
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 259
 LENGTH: 729
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-080-334-259

Query Match 69.1%; Score 94; DB 15; Length 729;
 Best Local Similarity 57.1%; Pred. No. 8.5e-05;
 Matches 16; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AGDMEISVKELRTILNRIISKHOLRT 28
 ||||| ||: ||: ||: ||: ||: ||:
 Db 570 AGDMEICADELKKVLNTVNVKHKOLKT 597

RESULT 12
 US-10-274-639-1
 Sequence 1, Application US/10274639
 Publication No. US20030232349A1
 GENERAL INFORMATION:
 APPLICANT: INCYTE GENOMICS, INC.
 APPLICANT: DELEGANE, Angelo M.; GANDHI, Ameena R.
 APPLICANT: HAFALIA, April J.A.; LU, Dyung Aina M.
 APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
 APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
 APPLICANT: NGUYEN, Danniell B.; LEE, Ernestine A.
 APPLICANT: KHAN, Farrah A.; YUE, Henry
 APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
 APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
 APPLICANT: YANG, Junming; THANGAVELU, Kavitha
 APPLICANT: DING, Li; KEARNEY, Liam
 APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.
 APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
 APPLICANT: BURFORD, Neil; WALIA, Narinder K.
 APPLICANT: LAL, Preeti G.; LEE, Sally
 APPLICANT: TODD, Stephen; LO, Terence P.
 APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
 APPLICANT: AZIMZAI, Yalda; LU, Yan
 TITLE OF INVENTION: PROTEASES
 FILE REFERENCE: PI-0167 USA
 CURRENT APPLICATION NUMBER: US/10/274,639
 CURRENT FILING DATE: 2002-10-18
 PRIOR APPLICATION NUMBER: PCT/US01/22397
 PRIOR FILLING DATE: 2001-07-17
 PRIOR APPLICATION NUMBER: US 60/220,063
 PRIOR FILLING DATE: 2000-07-21
 PRIOR APPLICATION NUMBER: US 60/221,680


```

; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/005,041
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 10/023,681
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/024,212
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 10/092,900
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 10/136,826
; PRIOR FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: 10/236,417
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 60/345,092
; PRIOR FILING DATE: 2002-01-04
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 122
; LENGTH: 773
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-336-472-122

Query Match      69.1%; Score 94; DB 15; Length 773;
Best Local Similarity 57.1%; Pred. No. 9.2e-05;
Matches 16; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY      1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db      614 AGDDMEICADELKKVLTIVVVKHKDLKT 641

Search completed: November 10, 2004, 16:36:02
Job time : 28.0062 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:36:51 ; Search time 5.66049 Seconds
(without alignments)
475.942 Million cell updates/sec

Title: US-10-092-750-18

Perfect score: 136

Sequence: 1 AGEDMEISVKELRTILNRISKKDLRT 28

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: P1r1.*

2: P1r2.*

3: P1r3.*

4: P1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136	100.0	714	1	CIHUH
2	121	89.0	320	2	A24815
3	117	86.0	585	2	S57195
4	115	84.6	705	1	CICTH
5	97	71.3	810	1	S57196
6	95	69.9	821	1	B34488
7	94	69.1	821	1	CIHUH3
8	91	66.9	184	2	A55143
9	91	66.9	263	2	A34466
10	91	66.9	266	1	CIPLG
11	91	66.9	266	1	CIREL
12	91	66.9	268	1	CIHUL
13	81	59.6	700	1	CIHUH2
14	81	59.6	700	2	S38361
15	80	58.8	422	2	B24815
16	80	58.8	700	2	S57194
17	68	50.0	703	2	A48764
18	57	41.9	339	2	D82366
19	56	41.2	675	1	KXWS
20	55.5	40.8	203	2	E72345
21	55	40.4	543	2	S73820
22	55	40.4	4077	2	T17484
23	53	39.0	1036	2	T05687
24	52	38.2	675	1	KXEOS
25	52	38.2	9376	2	T14593
26	51	37.5	3416	2	I59618
27	51	37.5	416	2	B72425
28	51	37.5	646	2	S38819
29	51	37.5	675	1	KXRTS

ALIGNMENTS

RESULT 1

CIHUH

calpain (EC 3.4.22.17) large chain 1 [validated] - human

N:Alternate names: calpain chain L-1; calpain I catalytic chain; low-calcium requiring, c

N:Contains: chemotactic factor

C:Species: Homo sapiens (man)

C>Date: 28-Dec-1987 #sequence, revision 28-Dec-1987 #text_change 09-Jul-2004

C:Accession: A26213; A36740; S10591

R:Aoki, K.; Imajoh, S.; Ohno, S.; Emori, Y.; Koike, M.; Kosaki, G.; Suzuki, K.

FEBS Lett. 205, 313-317, 1986

A:Title: Complete amino acid sequence of the large subunit of the low-Ca2+-requiring form

A:Reference number: A26213; MUID:86301172; PMID:3017764

A:Accession: A26213

A:Molecule type: mRNA

A:Residues: 1-714 <AOK>

A:Cross-references: UniProt:P07384; EMBL:X04366; NID:G29663; PIDN:CA27981.1; PID:G29664

R:Kumatsau, M.; Higashiyama, S.; Sato, K.; Onkubo, I.; Sasaki, M.

Biochem. Biophys. Res. Commun. 164, 875-882, 1989

A:Title: Calcium dependent cysteine proteinase is a precursor of a chemotactic factor for

A:Reference number: A36740; MUID:90056492; PMID:2554904

A:Accession: A36740

A:Molecule type: protein

A:Residues: 2-10 <KON>

A:Experimental source: erythrocytes

R:Scrimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saïdo, T.C.; Ohno, S.; Minami, Y.; S

Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990

A:Title: A novel member of the calcium-dependent cysteine protease family.

A:Reference number: S10589; MUID:90380278; PMID:2400579

A:Contents: annotation

A:Note: comparison with other gene products

C:Comment: Calpain I is activated by micromolar concentrations of calcium.

C:Genetics:

A:Gene: GDB:CAPN1; mu-CANP

A:Cross-references: GDB:119749; OMIM:114220

A:Map position: 11pter-11qter

C:Complex: heterodimer of L (large) and S (small) chains

C:Function:

A:Description: catalyzes the hydrolysis of peptides

A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and before

C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain

C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; duplication; EF1

P;2-714/Product: calpain large chain 1 #status predicted <MAT>

F;2-10/Product: chemotactic factor #status experimental <CHF>

F;85-337/Domain: calpain catalytic domain homology <CALP>

F;542-573/Domain: calmodulin repeat homology <EF1>

F;585-617/Domain: calmodulin repeat homology <EF2>

F;618-647/Domain: calmodulin repeat homology <EF3>

F;650-682/Domain: calmodulin repeat homology <EF4>

F;893-714/Domain: calmodulin repeat homology <EF5>

F;2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental

F;115,272,296/Active site: Cys, His, Asn #status predicted

```

Query Match      100.0%; Score 136; DB 1; Length 714;
Best Local Similarity 100.0%; Pred. No. 3.1e-10;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
DB 555 AGEDMEISVKELRTILNRIISKHKDLRT 582

RESULT 2
A24815
Calpain (EC 3.4.22.17) large chain 1 - rabbit (fragments)
N:Alternate names: calcium-activated neutral protease (CANP)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 24-Jan-1988 #sequence_revision 24-Feb-1994 #text_change 09-Jul-2004
C:Accession: A24815, A41418
R:Emori, Y.; Kawasaki, H.; Sugihara, H.; Imajoh, S.; Kawashima, S.; Suzuki, K.
J. Biol. Chem. 261, 9465-9471, 1986
A:Title: Isolation and sequence analyses of cDNA clones for the large subunits of two is
A:Reference number: A92594; MUID:86250902; PMID:2424911
A:Accession: A24815
A:Molecule type: mRNA
A:Residues: 19-320 <EMO>
A:Cross-references: UNIPROT:P06815; GB:M13363; NID:g165667; PIDN:AAA31456.1; PID:g165668
R:Kawasaki, H.; Imajoh, S.; Suzuki, K.
J. Biochem. 102, 393-400, 1987
A:Title: Separation of peptides on the basis of the difference in positive charge: simul
A:Reference number: A41418; MUID:89032960; PMID:3667575
A:Accession: A41418
A:Molecule type: protein
A:Residues: 1-18;125-154;313-320 <XAW>
A>Note: Sequence was deduced from composition by homology
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: blocked amino end; calcium binding; cysteine proteinase; duplication; EF hand
F:148-179/Domain: calmodulin repeat homology <EF1>
F:191-223/Domain: calmodulin repeat homology <EF2>
F:224-253/Domain: calmodulin repeat homology <EF3>
F:256-288/Domain: calmodulin repeat homology <EF4>
F:289-320/Domain: calmodulin repeat homology <EF5>

Query Match      89.0%; Score 121; DB 2; Length 320;
Best Local Similarity 85.7%; Pred. No. 1.4e-08;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
DB 161 AGEDLEISVRELQTLNRIISKHKDLRT 188

RESULT 3
S57195
Calpain (EC 3.4.22.17) large chain 1 - chicken (fragment)
N:Alternate names: mu-calpain heavy chain
C:Species: Gallus gallus (chicken)
C>Date: 28-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S57195
R:Sorimachi, H.; Tsukahara, T.; Okada-Ban, M.; Sugita, H.; Ishiura, S.; Suzuki, K.
Biochim. Biophys. Acta 1261, 381-393, 1995
A:Title: Identification of a third ubiquitous calpain species - chicken muscle expresses
A:Reference number: S57194; MUID:95260862; PMID:7742367
A:Accession: S57195
A:Molecule type: mRNA
A:Residues: 1-586 <SOR>
A:Cross-references: UNIPROT:O42133; EMBL:D38027; NID:c882070; PID:g882071
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: cysteine proteinase; EF hand; hydrolase
F:1-208/Domain: calpain catalytic domain homology (fragment) <CALP>
F:457-489/Domain: calmodulin repeat homology <EF1>

Query Match      86.0%; Score 117; DB 2; Length 586;
Best Local Similarity 85.7%; Pred. No. 8.6e-08;
Matches 24; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

Query Match      84.6%; Score 115; DB 1; Length 705;
Best Local Similarity 78.6%; Pred. No. 1.9e-07;
Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
DB 546 AGEDMEISVFLKTLNRIIVARHKDLRT 573

RESULT 5
S57196
Calpain (EC 3.4.22.17) large chain 3 - chicken
N:Alternate names: calpain p54 heavy chain; n-calpain-1 heavy chain
C:Species: Gallus gallus (chicken)
C>Date: 10-Sep-1995 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S57196
R:Sorimachi, H.; Tsukahara, T.; Okada-Ban, M.; Sugita, H.; Ishiura, S.; Suzuki, K.
Biochim. Biophys. Acta 1261, 381-393, 1995
A:Title: Identification of a third ubiquitous calpain species - chicken muscle expresses
A:Reference number: S57194; MUID:95260862; PMID:7742367
A:Accession: S57196
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-810 <SOR>
A:Cross-references: UNIPROT:Q92177; EMBL:D38028; NID:g882072; PIDN:BAA07230.1; PID:g15521
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: cysteine proteinase; EF hand; hydrolase
F:93-193/Domain: calpain catalytic domain homology <CALP>
F:746-778/Domain: calmodulin repeat homology <EFH>

```



```

A:Molecule type: mRNA
A:Residues: 44-445,'AA',448-458,'P',460-461,'P',463-484,'T',486-821 <SOR>
C:Genetics:
A:Gene: GDB:CAPN3; LGMD2A; LGMD2
A:A:Cross-references: GDB:119751; OMIM:114240; OMIM:253600
A:Map position: 15pter-15qter
A:Note: defects in this gene can cause limb girdle muscular dystrophy 2A
C:Complex: heterodimer of L (large) and S (small) chains
C:Function:
A:Description: catalyzes the hydrolysis of peptides
A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and before
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: calcium binding; cysteine proteinase; EF hand; hydrolase; muscular dystrophy
F:99-267,323-400/Domain: calpain catalytic domain homology #status atypical <CALP>
F:649-680/Domain: calmodulin repeat homology <EF1>
F:692-724/Domain: calmodulin repeat homology <EF2>
F:725-754/Domain: calmodulin repeat homology <EF3>
F:757-789/Domain: calmodulin repeat homology <EF4>
F:790-821/Domain: calmodulin repeat homology <EF5>
F:129,334,358/Active site: Cys, His, Asn #status predicted

Query Match 69.1%; Score 94; DB 1; Length 821;
Best Local Similarity 57.1%; Pred. No. 0.00014;
Matches 16; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
|||:||||| ||: ||| :|||:|||||:
Db 662 AGDDMEICADELKKVLTIVNKKHDLKT 689

RESULT 8
A55143
calpain (EC 3.4.22.17) light chain - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C:Accession: A55143
R:Graham-Stegenthaler, K.; Gauthier, S.; Davies, P.L.; Elce, J.S.
J. Biol. Chem. 269, 30457-30460, 1994
A:Title: Active recombinant rat calpain II. Bacterially produced large and small subunits
A:Reference number: A55143; MUID:95074051; PMID:7982961
A:Accession: A55143
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-184 <GRA>
A:Cross-references: UNIPROT:Q64537; GB:U10861; NID:G505657; PIDN:AAA64828.1; PID:G505658
C:Superfamily: calpain small chain; calmodulin repeat homology
C:Keywords: cysteine proteinase; EF hand; hydrolase
F:12-43/Domain: calmodulin repeat homology <EF1>
F:55-87/Domain: calmodulin repeat homology <EF2>
F:153-184/Domain: calmodulin repeat homology <EF3>

Query Match 66.9%; Score 91; DB 2; Length 184;
Best Local Similarity 53.6%; Pred. No. 7.6e-05;
Matches 15; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
|||:||||| ||: ||| :|||:|||||:
Db 25 AGDDMEVSATELMNILNKVTRHPDLKT 52

RESULT 9
A34466
calpain (EC 3.4.22.17) light chain - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 09-Jul-2004
C:Accession: A34466
R:McLelland, P.; Lash, J.A.; Hathaway, D.R.
J. Biol. Chem. 264, 17428-17431, 1989
A:Title: Identification of major autolytic cleavage sites in the regulatory subunit of va
A:Reference number: A34466; MUID:90008911; PMID:2551592
A:Accession: A34466
A>Status: preliminary

```


DB 109 AGDDMEVSNATLMLNKLNVTRHPDLKT 136

RESULT 13
CIHUH2

N:calpain (EC 3.4.22.17) large chain 2 - human
C:Alternate names: calpain chain I-2; calpain II catalytic chain; high-calcium requiring
C:Species: Homo sapiens (man)
C>Date: 21-Nov-1993 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C/Accession: S10590; A3218; A33529
R:Sorimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saïdo, T.C.; Ohno, S.; Minami, Y.; S
Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990
A>Title: A novel member of the calcium-dependent cysteine protease family.
A/Reference number: S10589; MUID:90380278; PMID:2400579
A/Accession: S10590
A/Molecule type: mRNA
A/Residues: 1-700 <SOR>
A/Cross-references: UNIPROT:P17655
R:Imajoh, S.; Ackl, K.; Ohno, S.; Emori, Y.; Kawasaki, H.; Sugihara, H.; Suzuki, K.
Biochemistry 27, 8122-8128, 1988
A>Title: Molecular cloning of the cDNA for the large subunit of the high-Ca(2+)-requir
A/Reference number: A31218; MUID:89166474; PMID:2852952
A/Accession: A31218
A/Molecule type: mRNA; protein
A/Residues: 1-210,'I','212-394,'D','396-445','I','447-700 <IMA>
A/Cross-references: GB:M23254; NID:G511636; PIDN:AAA35645.1; PID:G511637
A>Note: Parts of this sequence were determined by protein sequencing; the amino end of t
R:Hata, A.; Ohno, S.; Akita, Y.; Suzuki, K.
J. Biol. Chem. 264, 6404-6411, 1989
A>Title: Tandemly reiterated negative enhancer-like elements regulate transcription of a
A/Reference number: A33529; MUID:89197947; PMID:2539381
A/Accession: A33529
A/Molecule type: DNA
A/Residues: 1-67,'G','69-72','IE','75-78','R' <HAT>
A/Cross-references: DDBJ:J04700; NID:G179910; PIDN:AAA52760.1; PID:G463086
C/Genetics:
A/Gene: GDB:CAPN2; mCANP; CANPml
A/Cross-references: GDB:I19750; OMIM:114230
A/Map position: lpter-lqter
C/Complex: heterodimer of L (large) and S (small) chains
C/Function:
A/Description: catalyzes the hydrolysis of peptides
A/Note: cleaves preferentially after tyrosine, methionine, or arginine residues and bef
C/Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C/Keywords: acetylated amino end; calcium binding; cysteine proteinase; EF hand; hetero
F:2-700/Product: calpain large chain 2 #status predicted <MAT>
F:75-327/Domain: calpain catalytic domain homology <CALP>
F:529-560/Domain: calmodulin repeat homology <BF1>
F:572-604/Domain: calmodulin repeat homology <BF2>
F:605-634/Domain: calmodulin repeat homology <BF3>
F:637-669/Domain: calmodulin repeat homology <BF4>
F:7/Modified site: acetylated amino end (Ala) (in mature form) #status predicted
F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 59.6%; Score 81; DB 1; Length 700;
Best Local Similarity 53.6%; Pred.No. 0.0063;
Matches 15; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

OY 1 AGEDEISVKELRTILNLRIISKHKDLRT 28
DB 542 AGDAEISAPELOTLIRVLAKQDIKS 569
|||||
|||

RESULT 14
S38361

N:calpain (EC 3.4.22.17) II large chain - rat
C:Alternate names: calpain II 80K chain
C/Species: Rattus norvegicus (Norway rat)
C>Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 09-Jul-2004
C/Accession: S38361; S08650; S39751
R:DeLuca, C.I.; Davies, P.L.; Samis, J.A.; Elce, J.S.

A:Reference number: S38361; MUID:94032492; PMID:8218419
A:Accession: S38361
A:Molecule type: mRNA
A:Residues: 1-700
A:Cross-references: UNIPROT:Q07009; EMBL:LC0120; NID:G402665; PIDN:AAA16327.1; PID:G40266
R:Samis, J.A.; Back, D.W.; Graham, S.J.; Elice, J.S.
submitted to the EMBL Data Library, February 1990
A:Reference number: S08650
A:Accession: S08650
A:Molecule type: DNA
A:Residues: 380-439 <SAM>
A:Cross-references: EMBL:A51772
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer; hydri
F:75-327/Domain: calpain catalytic domain homology <CALP>
F:539-560/Domain: calmodulin repeat homology <EF1>
F:572-604/Domain: calmodulin repeat homology <EF2>
F:605-634/Domain: calmodulin repeat homology <EF3>
F:637-669/Domain: calmodulin repeat homology <EF4>
F:105,262,286/Active site: Cys, His, Asn #status Predicted

Query Match 59.6%; Score 81; DB 2; Length 700;
Best Local Similarity 53.6%; Pred. No. 0.0063;
Matches 15; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHDLRT 28
||||| ||||| ||||| ||||| ||||| |||||
Db 542 AGEDAEISAFELQILRLVLAKREDIKS 569

RESULT 15
B24815
calpain (EC 3.4.22.17) large chain 2 - rabbit (fragment)
N:Alternate names: calcium-activated neutral proteinase (CANP)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 24-Jan-1988 #sequence_revision 24-Jan-1988 #text_change 09-Jul-2004
C:Accession: B24815
R:Emori, Y.; Kawasaki, H.; Sugihara, H.; Imajoh, S.; Kawashima, S.; Suzuki, K.
J. Biol. Chem. 261, 9465-9471, 1986
A:Title: Isolation and sequence analyses of cDNA clones for the large subunits of two iso
A:Reference number: A92594; MUID:86250902; PMID:2424911
A:Accession: B24815
A:Molecule type: mRNA
A:Residues: 1-422 <EMO>
A:Cross-references: UNIPROT:P06814; GB:M13797; NID:G165665; PIDN:AAA31455.1; PID:G165666
C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
C:Keywords: calcium binding; cysteine proteinase; EF hand; hydrolase
F:251-282/Domain: calmodulin repeat homology <EF1>
F:294-326/Domain: calmodulin repeat homology <EF2>
F:327-356/Domain: calmodulin repeat homology <EF3>
F:359-391/Domain: calmodulin repeat homology <EF4>

Query Match 58.8%; Score 80; DB 2; Length 422;
Best Local Similarity 53.6%; Pred. No. 0.0052;
Matches 15; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHDLRT 28
||||| ||||| ||||| ||||| ||||| |||||
Db 264 AGEDAEISAFELQILRLVLAKRQDIKT 291

Search completed: November 10, 2004, 12:29:22
Job time : 6.66049 secs

Search completed: November 10, 2004, 12:29:22
Job time : 6.66049 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01 ; Search time 30.6358 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-18

Perfect score: 136

Sequence: 1 AGEDMEISVKELRTILNRIISKHDLRT 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	136	100.0	206	2 Q9GL55	Q9GL55 oviv aries
2	136	100.0	714	1 CAN1_HUMAN	P07384 homo sapien
3	136	100.0	714	1 CAN1_MACFA	Q9GL52 macaca fasc
4	136	100.0	714	2 Q6DHV4	Q6DHV4 homo sapien
5	136	100.0	716	1 CAN1_BOVIN	Q27970 bos taurus
6	136	100.0	716	2 Q9N185	Q9N185 bos taurus
7	133	97.8	650	2 Q9N0M7	Q9N0M7 sus scrofa
8	133	97.8	714	1 CAN1_PIG	F35750 sus scrofa
9	129	94.9	713	1 CAN1_MOUSE	O35350 mus musculus
10	129	94.9	713	1 CAN1_RAT	P97571 rattus norv
11	129	94.9	713	2 Q80Y25	Q80Y25 mus musculus
12	129	94.9	713	2 Q8C2J1	Q8C2J1 mus musculus
13	129	94.9	713	2 AAH61880	AAH61880 rattus no
14	121	89.0	302	1 CAN1_RABIT	P06815 oryctolagus
15	119	87.5	704	2 Q7SFA9	Q7SFA9 brachydanio
16	117	86.0	715	2 Q42133	Q42133 gallus gall
17	115	84.6	705	1 CANX_CHICK	P00789 gallus gall
18	115	84.6	705	2 Q9YTC1	Q9YTC1 coturnix co
19	114	83.8	704	2 Q6PAF2	Q6PAF2 xenopus lae
20	114	83.8	704	2 Q7ZUR1	Q7ZUR1 brachydanio
21	114	83.8	704	2 AAH60341	AAH60341 xenopus l
22	114	83.8	704	2 Q8UW96	Q8UW96 xenopus lae
23	114	83.8	724	2 Q7ZXQ4	Q7ZXQ4 xenopus lae
24	107	78.7	713	2 Q6QDT7	Q6QDT7 bos taurus
25	103	75.7	41	2 Q6QDT7	Q6QDT7 bos taurus
26	103	75.7	41	2 AAS66957	AAS66957 bos tauru
27	101	74.3	709	2 Q9XSJ1	Q9XSJ1 bos taurus
28	101	74.3	709	2 Q9XSJ2	Q9XSJ2 sus scrofa
29	101	74.3	821	1 CAN3_PIG	F43368 sus scrofa
30	101	74.3	821	1 CAN3_BOVIN	P51186 bos taurus
31	101	74.3	822	1 CAN3_SHEEP	Q9Tth8 oviv aries

```

32 97 71.3 810 1 CAN3_CHICK Q92177 gallus gall
33 95 69.9 191 2 Q8C2H4 Q8C2H4 mus musculus
34 95 69.9 652 2 Q762C5 Q762C5 mus musculus
35 95 69.9 652 2 BAD16652 BAD16652 mus muscu
36 95 69.9 709 2 O88702 O88702 rattus norv
37 95 69.9 709 2 O88977 O88977 mus musculus
38 95 69.9 757 2 O70482 O70482 rattus norv
39 95 69.9 821 1 CAN3_MOUSE Q64691 mus musculus
40 95 69.9 821 1 CAN3_RAT P16259 rattus norv
41 94 69.1 230 2 Q7L4R0 Q7L4R0 homo sapien
42 94 69.1 230 2 AAH07810 AAH07810 homo sapi
43 94 69.1 728 2 Q762C8 Q762C8 homo sapien
44 94 69.1 728 2 BAD16649 BAD16649 homo sapi
45 94 69.1 815 1 CAN3_MACFA Q95197 macaca fasc

```

ALIGNMENTS

```

RESULT 1
O9GL55 PRELIMINARY; PRT; 206 AA.
AC Q9GL55;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Calpain 1 (Fragment)
OS Ovis aries (Sheep)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Chung H.Y., Davis M.E., Hines H.C.;
RL Submitted (OC1-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF316574; AAG31809.1; -
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0004198; P:calpain activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010983; EF_Hand like.
DR InterPro; IPR001300; Peptidase_C2.
DR Pfam; PF01067; Calpain_III; 1.
DR SMART; SM00720; calpain_III; 1.
FT NON_TER 1
FT NON_TER 206
SQ SEQUENCE 206 AA; 23837 MW; 5E7F69C4133043FF CRC64;

Query Match 100.0%; Score 136; DB 2; Length 206;
Best Local Similarity 100.0%; Pred.No. 7.3e-10;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 AGEDMEISVKELRTILNRIISKHDLRT 28
|||||
Db 169 AGEDMEISVKELRTILNRIISKHDLRT 196
|||||

```

```

RESULT 2
CAN1_HUMAN STANDARD; PRT; 714 AA.
ID CAN1_HUMAN
AC P07384;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Calpain 1, large [catalytic] subunit (EC 3.4.22.52) (Calcium-activated
DE neutral proteinase) (CANP) (Mu-type) (muCANP) (Micromolar-calpain).
GN Name=CAPN1; Synonyms=CAPN1;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=86301172; PubMed=3017764;
 RA Aoki K., Imajoh S., Ohno S., Emori Y., Koike M., Kosaki G., Suzuki K.;
 RT "Complete amino acid sequence of the large subunit of the low-Ca²⁺-
 RT requiring form of human Ca²⁺-activated neutral protease (muCAMP)
 RL deduced from its cDNA sequence."; PubMed 205:313-317(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90380278; PubMed=2400579;
 RA Sorimachi H., Ohmi S., Emori Y., Kawasaki H., Saido T.C., Ohno S.,
 RA Minami Y., Suzuki K.;
 RT "A novel member of the calcium-dependent cysteine protease family.";
 RL Biol. Chem. Hoppe-Seyler 371:171-176(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney, and Placenta;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Vax S.J., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A.C., Rodrigues S., Sanchez A.,
 RA Whitting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyzes limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -!- CATALYTIC ACTIVITY: Broad endopeptidase specificity.
 CC -!- COFACTOR: Binds 3 calcium ions.
 CC -!- ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin.
 CC -!- SUBUNIT: Forms a heterodimer with a small (regulatory) subunit
 CC (CAPN1).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca(2+) binding (By similarity).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- SIMILARITY: Belongs to peptidase family C2.
 CC -!- SIMILARITY: Contains 1 Calpain catalytic domain.
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -!- DATABASE: NAME=CaBP; NOTE=Calpain;
 CC WWW="http://structbio.vanderbilt.edu/cabp_database/general/prot_pages/calpain.h
 CC -!- DATABASE: NAME=Calpains homepage;
 CC WWW="http://ag.arizona.edu/calpains".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X04366; CAA27881.1; -;
 CC EMBL; BC008751; AAH08751.1; -;
 CC EMBL; BC017200; AAH17200.1; -;
 CC PIR; A26213; C1HUH.
 CC HSP; P97571; 1KXR.
 CC MEROPS; C02.001; -;
 CC Genew; HGNC:1476; CAPN1.
 CC MIM; 114220; -;

DR GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR010983; EF-Hand-like.
 DR InterPro; IPR001300; Peptidase_C2.
 DR InterPro; IPR001169; Pept_cys_acsite.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF06448; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN_III; 1.
 DR SMART; SM00720; calpain_III; 1.
 DR SMART; SM00230; Cyspc; 1.
 DR SMART; SM00054; EFh; 3.
 DR PROSITE; PS0203; CALPAIN_CAT; 1.
 DR PROSITE; PS0018; EF_HAND; 2.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00839; THIOL_PROTEASE_HIS; FALSE_NEG.
 KW Calcium-binding; Hydrolase; Multigene family; Thiol protease.
 FT DOMAIN 55 354 Calpain catalytic.
 FT DOMAIN 355 526 Domain III.
 FT DOMAIN 527 542 Linker.
 FT DOMAIN 543 713 Domain IV.
 FT DOMAIN 554 565 Ancestral calcium site 1 (Potential).
 FT CA_BIND 565 609 EF-hand 2.
 FT CA_BIND 609 639 EF-hand 3.
 FT DOMAIN 639 674 Ancestral calcium site 4 (Potential).
 FT DOMAIN 674 704 Ancestral calcium site 5 (Potential).
 FT ACT_SITE 115 115 By similarity.
 FT ACT_SITE 272 272 By similarity.
 FT ACT_SITE 296 296 By similarity.
 FT CONFLICT 548 548 K -> N (in Ref. 3; AAH08751).
 SQ SEQUENCE 714 AA; 81889 MW; 1CB6D7C56D063498 CRC64;
 Query Match 100.0%; Score 136; DB 1; Length 714;
 Best Local Similarity 100.0%; Pred. No. 2.5e-09;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGEDMEISVKELRTILNRIISKHDLRT 28
 DB 555 AGEDMEISVKELRTILNRIISKHDLRT 582
 RESULT 3
 CANL_MACFA STANDARD; PRT; 714 AA.
 ID CANL_MACFA
 AC Q9GUG2;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Calpain 1, large (catalytic) subunit (EC 3.4.22.52) (Calcium-activated
 DE neutral proteinase) (CAMP) (Mu-type) (muCAMP) (Micromolar-calpain).
 GN Name=CAPN1; Synonyms=CANPL1;
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OC NCBI_TaxID=9541;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RX MEDLINE=21299050; PubMed=11406271; DOI=10.1016/S0167-4781(01)00212-3;
 RA Nakajima T., Fukui C., Azuma M., Ma H., Shearer T.R.;
 RT "Different expression patterns for ubiquitous calpains and Capn3
 RT splice variants in monkey ocular tissues";
 RL Biochim. Biophys. Acta 1519:55-64(2001).
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyzes limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction (By similarity).
 CC -!- CATALYTIC ACTIVITY: Broad endopeptidase specificity.
 CC -!- COFACTOR: Binds 3 calcium ions (By similarity).
 CC -!- ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin (By similarity).
 CC -!- SUBUNIT: Forms a heterodimer with a small (regulatory) subunit

```

CC CC (-CAPNS1).
CC CC SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC CC membrane upon Ca++ binding (By similarity).
CC CC TISSUE SPECIFICITY: Ubiquitous.
CC CC SIMILARITY: Belongs to peptidase family C2.
CC CC SIMILARITY: Contains 1 Calpain catalytic domain.
CC CC SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by, and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; AF284440; AAG22770.1; -.
CC CC HSPG; P97571; 1KXR.
CC CC MEROPS; C02.001; -.
CC CC InterPro; IPR002048; EF-hand.
CC CC InterPro; IPR010983; EF_Hand_like.
CC CC InterPro; IPR001300; Peptidase C2.
CC CC InterPro; IPR00169; Pept_cys_acsite.
CC CC Pfam; PF01067; Calpain_III; 1.
CC CC Pfam; PF00036; ehand; 3.
CC CC Pfam; PF00648; Peptidase_C2; 1.
CC CC PRINTS; PR00704; CALPAIN_III; 1.
CC CC SMART; SM00720; calpain_III; 1.
CC CC SMART; SM00230; Cyspc; 1.
CC CC SMART; SM00054; EFh; 3.
CC CC PROSITE; PS02023; CALPAIN_CAT; 1.
CC CC PROSITE; PS00018; EF_HAND; 2.
CC CC PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
CC CC PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
CC CC PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
CC CC Calcium-binding; Hydroxylase; Multigene family; Thiol protease.
CC CC DOMAIN 55 354
CC CC DOMAIN 355 526
CC CC DOMAIN 527 542
CC CC DOMAIN 543 713
CC CC DOMAIN 554 565
CC CC DOMAIN 598 609
CC CC CA_BIND 628 639
CC CC CA_BIND 663 674
CC CC DOMAIN 693 704
CC CC ACT_SITE 115 115
CC CC ACT_SITE 272 272
CC CC ACT_SITE 296 296
CC CC SEQUENCE 714 AA; 81849 MW; 834690C214D57AD8 CRC64;

Query Match 100.0%; Score 136; DB 1; Length 714;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHDLRT 28
Db 555 AGEDMEISVKELRTILNRIISKHDLRT 582

RESULT 4
Q6DHV4 PRELIMINARY; PRT; 714 AA.
ID Q6DHV4
AC Q6DHV4;
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DB CAPN1 protein.
GN Name=CAPN1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner J., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Halton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Gimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX Strausberg R.;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC075862; AAH75862.1; -.
SQ SEQUENCE 714 AA; 81830 MW; 46B6556D4707428F CRC64;

Query Match 100.0%; Score 136; DB 2; Length 714;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGEDMEISVKELRTILNRIISKHDLRT 28
Db 555 AGEDMEISVKELRTILNRIISKHDLRT 582

RESULT 5
CAN1_BOVIN STANDARD; PRT; 716 AA.
ID CAN1_BOVIN
AC Q27970; Q9NOU3; Q9NOV6;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-OCT-2004 (rel. 45, Last sequence update)
DT 01-OCT-2004 (rel. 45, Last annotation update)
DE Calpain 1, large [catalytic] subunit (BC 3.4.22.52) (Calcium-activated
DE neutral proteinase) (CAMP) (Mu-type) (mucANP) (Micromolar-calpain).
GN Name=CAPN1;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20500419; PubMed=11048924;
RA Smith T.P.L., Casas E., Rexroad C.E. III, Kappes S.M., Keele J.W.;
RT "Bovine CAPN1 maps to a region of BTA29 containing a quantitative
RT trait locus for meat tenderness.";
RL J. Anim. Sci. 78:2589-2594(2000).
RN [2]
RP SEQUENCE OF 530-625 FROM N.A.
RC TISSUE=Skeletal muscle;
RA Sun W., Bidwell C.A., Ji S., Hancock D.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction.
CC -!- CATALYTIC ACTIVITY: Broad endopeptidase specificity.
CC -!- COFACTOR: Binds 3 calcium ions.
CC -!- ENZYME REGULATION: Activated by micromolar concentrations of

```

CC calcium and inhibited by calpastatin.
CC -!- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family C2.
CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshifts in positions 582 and 617.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF248054; AAF64504.2; -.
DR EMBL; AF252504; AAF64504.2; JOINED.
DR EMBL; U07849; AAA18454.1; ALT_FRAME.
DR PIR; S16181; S16181.
DR HSSP; P17655; 1KXU.
DR MEROPS; C02.001; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001300; Peptidase_C2.
DR InterPro; IPR00169; Pept_cys_acsite.
DR Pfam; PF01067; Calpain_III; 1.
DR Pfam; PF00036; ehand; 3.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN_III; 1.
DR SMART; SM00720; calpain_III; 1.
DR SMART; SM00230; Cyspc; 1.
DR SMART; SM00054; EFh; 3.
DR PROSITE; PS0203; CALPAIN CAT; 1.
DR PROSITE; PS00018; EF HAND; 2.
DR PROSITE; PS00540; THIOL_PROTEASE_CYS; 1.
DR PROSITE; PS00139; THIOL_PROTEASE_HIS; FALSE_NEG.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
KW Calcium-binding; Hydroxylase; Multigene family; Thiol protease.
FT DOMAIN 55 354
FT DOMAIN 355 528
FT DOMAIN 529 544
FT DOMAIN 545 715
FT DOMAIN 556 567
FT CA_BIND 600 611
FT CA_BIND 630 641
FT DOMAIN 665 676
FT DOMAIN 695 706
FT ACT_SITE 115 115
FT ACT_SITE 272 272
FT ACT_SITE 296 296
FT CONFLICT 530 530
FT CONFLICT 624 624
SQ SEQUENCE 716 AA; 82249 MW; 8FFD61331DAB8A5 CRC64;

Query Match 100.0%; Score 136; DB 1; Length 716;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
DB 557 AGEDMEISVKELRTILNRIISKHKDLRT 584

RESULT 6
Q9N185 PRELIMINARY; PRT; 716 AA.
AC Q9N185;
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Micromolar calcium-dependent neutral protease large subunit.

Query Match 100.0%; Score 136; DB 1; Length 716;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
DB 557 AGEDMEISVKELRTILNRIISKHKDLRT 584

RESULT 6
Q9N185 PRELIMINARY; PRT; 716 AA.
AC Q9N185;
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Micromolar calcium-dependent neutral protease large subunit.

GN Name=CAPN1;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith T.P.L., Casas E., Rexroad C.E., Kappes S.M., Keele J.W.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP221129; AAF32364.1; -.
DR PIR; S16181; S16181.
DR HSSP; P97571; 1KXU.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004198; F:calpain activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR010983; EF_Hand_like.
DR InterPro; IPR001300; Peptidase_C2.
DR InterPro; IPR00169; Pept_cys_acsite.
DR Pfam; PF01067; Calpain_III; 1.
DR Pfam; PF00036; ehand; 3.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN_III; 1.
DR SMART; SM00720; calpain_III; 1.
DR SMART; SM00230; Cyspc; 1.
DR SMART; SM00054; EFh; 3.
DR PROSITE; PS0203; CALPAIN CAT; 1.
DR PROSITE; PS00018; EF HAND; UNKNOWN 2.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
KW Protease.
SQ SEQUENCE 716 AA; 82207 MW; A7F55C197BD5DF23 CRC64;

Query Match 100.0%; Score 136; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
DB 557 AGEDMEISVKELRTILNRIISKHKDLRT 584

RESULT 7
Q9N0M7 PRELIMINARY; PRT; 650 AA.
AC Q9N0M7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Micromolar calcium-activated neutral protease 1 isoform B.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith T.P.L., Simmen F.A., Zhao G., Vallet J.L.;
RL "Nucleotide sequences of two isoforms of porcine micromolar calcium-
RT activated neutral protease 1 cDNA".;
RL J. Anim. Sci. 79:552-553(2001).
DR EMBL; AF263609; AAF73443.1; -.
DR HSSP; P97571; 1KXU.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004198; F:calpain activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR010983; EF_Hand_like.
DR InterPro; IPR001300; Peptidase_C2.

DR InterPro; IPR000169; Pept_cys_acsite.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00036; ehand; 2.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN_III; 1.
 DR SMART; SM00720; calpain_III; 1.
 DR SMART; SM00230; CysPC; 1.
 DR SMART; SM00054; Efn; 2.
 DR PROSITE; PS0203; CALPAIN_CAT; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 2.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 KW Protease.
 SQ SEQUENCE 650 AA; 74033 MW; 2EF93FF569D90FD7 CRC64;
 Query Match 97.8%; Score 133; DB 2; Length 650;
 Best Local Similarity 96.4%; Pred. No. 5.7e-09;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGEDMEISVKELRLNRIISKHDLRT 28
 |||||.....
 Db 555 AGEDMEISVKELRLNRIISKHDLRT 582
 |||||.....
 RESULT 8
 CANI_PIG STANDARD; PRT; 714 AA.
 AC P35750; Q29600; O9XOM6;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.52) (Calcium-activated
 DE neutral proteinase) (CAMP) (Mu-type) (mucAMP) (Micromolar-calpain).
 GN Name=CAPN1;
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21087428; PubMed=11219468;
 RA Smith T.P.L., Simmen F.A., Zhao G., Vallet J.L.;
 RT "Rapid communication: nucleotide sequences of two isoforms of porcine
 RT micromolar calcium-activated neutral protease 1 cDNA.";
 RL J. Anim. Sci. 79:552-553(2001).
 RN [2]
 RP SEQUENCE OF 326-415 FROM N.A.
 RC TISSUE=Small intestine;
 RA Winteroe A.K., Fredholm M., Davies W.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 528-623 FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=94146155; PubMed=8312396;
 RA Sun W., Ji S.Q., Ebert P.J., Bidwell C.A., Hancock D.L.;
 RT "Cloning the partial cDNAs of mu-calpain and m-calpain from porcine
 RT skeletal muscle.";
 RL Biochimie 75:931-936(1993).
 CC -!- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -!- CATALYTIC ACTIVITY: Broad endopeptidase specificity.
 CC -!- COFACTOR: Binds 3 calcium ions (By similarity).
 CC -!- ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin.
 CC -!- SUBUNIT: Forms a heterodimer with a small (regulatory) subunit
 CC (CAPNS1).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca(2+) binding (By similarity).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- SIMILARITY: Belongs to peptidase family C2.
 CC -!- SIMILARITY: Contains 1 Calpain catalytic domain.
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

 DR EMBL; AF263610; AAF73444.1; -
 DR EMBL; F14611; CAA23154.1; -
 DR EMBL; U01180; AAG65125.1; -
 DR HSSP; P97571; 1KX8.
 DR MEROPS; C02.001; -
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR010983; EF_Hand_like.
 DR InterPro; IPR001300; Peptidase_C2.
 DR InterPro; IPR000169; Pept_cys_acsite.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN_CAT; 1.
 DR PROSITE; PS0203; CALPAIN_CAT; 1.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS00640; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOI_PROTEASE_HIS; FALSE NEG.
 KW Calcium-binding; Hydrolase; Multigene family; Thiol protease.
 FT DOMAIN 55 354
 FT Domain III.
 FT DOMAIN 355 526
 FT Domain IV.
 FT DOMAIN 527 542
 FT DOMAIN 543 713
 FT DOMAIN 554 565
 FT CA_BIND 598 609
 FT CA_BIND 628 639
 FT DOMAIN 663 674
 FT DOMAIN 693 704
 FT ACT_SITE 115 115
 FT ACT_SITE 272 272
 FT ACT_SITE 296 296
 FT ACT_SITE 528 528
 FT CONFLICT 531 531
 FT CONFLICT 541 541
 FT CONFLICT 622 622
 SQ SEQUENCE 714 AA; 81738 MW; 0BB31DE4FC56363A CRC64;
 Query Match 97.8%; Score 133; DB 1; Length 714;
 Best Local Similarity 96.4%; Pred. No. 6.2e-09;
 Matches 27; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGEDMEISVKELRLNRIISKHDLRT 28
 |||||.....
 Db 555 AGEDMEISVKELRLNRIISKHDLRT 582
 |||||.....
 RESULT 9
 CANI_MOUSE STANDARD; PRT; 713 AA.
 AC O35350; O88666;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.52) (Calcium-activated
 DE neutral proteinase) (CAMP) (Mu-type) (mucAMP) (Micromolar-calpain).
 GN Name=Capn1; Synonyms=Capn1, Capal;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Cf-1;
 RA Poirier C., Poussard S., Faust D.M., Imaizumi-Sherer T., Weiss M.C.,
 RA Ducastaing A., Montarras D., Pinset C., Guenet J.-L.;

[illegible]

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U53858; AAC53001.1; --
 DR PDB; 1KXR; X-ray; A/B=26-364.
 DR MR0PS; C02.001; --
 DR RGD; 2267; Capni.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR010983; EF_Hand_Like.
 DR InterPro; IPR001300; Peptidase_C2.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00036; ehand_3; 1.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR PRINTS; PR00704; CALPAIN_III; 1.
 DR SMART; SM00720; calpain_III; 1.
 DR SMART; SM00230; CysPc; 1.
 DR SMART; SM00054; EFh; 3.
 DR PROSITE; PS02023; CALPAIN_CAT; 1.
 DR PROSITE; PS00018; EF_HAND_2.
 DR PROSITE; PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
 DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
 KW 3D-structure; Calcium-binding; Hydrolase; Thiol protease.
 FT DOMAIN 55 354
 FT DOMAIN 355 525
 FT DOMAIN 526 541
 FT DOMAIN 542 712
 FT DOMAIN 553 564
 FT CA_BIND 597 608
 FT CA_BIND 627 638
 FT DOMAIN 662 673
 FT DOMAIN 692 703
 FT ACT_SITE 115 115
 FT ACT_SITE 272 272
 FT ACT_SITE 296 296
 FT STRAND 35 36
 FT HELIX 37 39
 FT HELIX 42 51
 FT TURN 52 53
 FT TURN 59 60
 FT HELIX 65 68
 FT TURN 72 73
 FT TURN 75 76
 FT HELIX 81 82
 FT STRAND 84 86
 FT HELIX 88 91
 FT STRAND 96 96
 FT HELIX 104 106
 FT STRAND 108 108
 FT HELIX 115 124
 FT TURN 125 126
 FT HELIX 128 134
 FT TURN 143 144
 FT STRAND 148 155
 FT TURN 156 157
 FT STRAND 158 165
 FT STRAND 168 171
 FT TURN 172 173
 FT STRAND 174 175
 FT STRAND 179 179
 FT TURN 183 184
 FT STRAND 185 185
 FT HELIX 187 198
 FT TURN 199 200
 FT HELIX 203 205
 FT HELIX 210 218
 FT STRAND 221 226
 FT HELIX 227 229
 FT TURN 232 233
 FT HELIX 234 244
 FT TURN 245 245

FT STRAND 247 251
 FT HELIX 257 259
 FT TURN 260 261
 FT STRAND 263 263
 FT TURN 265 266
 FT STRAND 269 269
 FT TURN 270 271
 FT STRAND 274 283
 FT STRAND 288 295
 FT TURN 297 298
 FT TURN 306 307
 FT TURN 309 310
 FT HELIX 312 316
 FT STRAND 319 325
 FT STRAND 333 337
 FT HELIX 338 344
 FT STRAND 347 352
 SQ SEQUENCE 713 AA; 82118 MW; 6E664600B0EFAEBB CRC64;
 Query Match 94.9%; Score 129; DB 1; Length 713;
 Best Local Similarity 92.9%; Pred.No.2.1e-08;
 Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGEDMEISVKELTILNRIISKHKDLRT 28
 Db 554 AGDDMEISVKELTILNRIISKHKDLRT 581
 RESULT 11
 ID Q80Y25 PRELIMINARY; PRT; 713 AA.
 AC Q80Y25;
 DT 01-JUN-2003 (TRENBLrel. 24, Created)
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Capn1 protein.
 GN Name=Capn1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Salivary gland;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Salivary gland;
 RA Klausner R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC050276; AAH50276.1; --
 DR HSSP; P97571; 1KXR.
 DR MGD; MGI:86263; Capn1.

DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0004199; F:calpain activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR010983; EF-hand like.
 DR InterPro; IPR001300; Peptidase C2.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00036; ehand; 3.
 DR PRINTS; PD00704; CALPAIN.
 DR PRODOM; PD000012; EF-hand; 1.
 DR SMART; SM00720; calpain_III; 1.
 DR SMART; SM00230; Cyspc; 1.
 DR SMART; SM00054; Eph; 3.
 DR PROSITE; PS0203; CALPAIN_CAT; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 2.
 DR PROSITE; PS00139; THIOLE PROTEASE_CYS; 1.
 SQ SEQUENCE 713 AA; 82134 MW; 2EFF1719B935B860 CRC64;

Query Match 94.9%; Score 129; DB 2; Length 713;
 Best Local Similarity 92.9%; Pred. No. 2.1e-08;
 Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHDLRT 28
 DB 554 AGDDMEISVKELQILNRIISKHDLRT 581

RESULT 12
 Q8C2J1 PRELIMINARY; PRT; 713 AA.
 AC Q8C2J1
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:E430019P14 product:calpain 1, full insert sequence.
 GN Name=Capn1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.,
 RT "High-efficiency full-length cDNA cloning";
 RL Meth. Enzymol. 303:19-44 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 403:685-690 (2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RA The FANTOM Consortium,
 RT the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=20493174; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes."; Genome Res. 10:1617-1630 (2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RX MEDLINE=20531013; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer";
 RL Genom. Res. 10:1757-1771 (2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOD; TISSUE=Thymus;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P., Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T., Horii F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y., Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H., Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M., Tagawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T., Tomaru A., Toya I., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AX088547; BAC40416.1; --
 DR HSSP; P97571; 1KXR.
 DR MGD; MGI:88263; Capn1.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0004199; F:calpain activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR010983; EF-hand like.
 DR InterPro; IPR001300; Peptidase C2.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00036; ehand; 3.
 DR PRINTS; PD00704; CALPAIN.
 DR PRODOM; PD000012; EF-hand; 1.
 DR SMART; SM00720; calpain_III; 1.
 DR SMART; SM00230; Cyspc; 1.
 DR SMART; SM00054; Eph; 3.
 DR PROSITE; PS0203; CALPAIN_CAT; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 2.
 DR PROSITE; PS00139; THIOLE PROTEASE_CYS; 1.
 SQ SEQUENCE 713 AA; 82152 MW; 3E1FCB4D5802B864 CRC64;

Query Match 94.9%; Score 129; DB 2; Length 713;
 Best Local Similarity 92.9%; Pred. No. 2.1e-08;
 Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHDLRT 28
 DB 554 AGDDMEISVKELQILNRIISKHDLRT 581

RESULT 13
 AAH61880 PRELIMINARY; PRT; 713 AA.
 ID AAH61880
 AC AAH61880
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE Calpain 1.

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T.J., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skalska U., Smalls J., Schermer A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RA Strausberg R.;
RA EMBL; BC061880; AAH61880.1; -
SQ SEQUENCE 713 AA; 82119 MW; 6B66460B0EFAEBB CRC64;

Query Match 94.9%; Score 129; DB 2; Length 713;
Best Local Similarity 92.9%; Pred. No. 2;le-08;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db 554 AGDDMEISVKELQILNRIISKHKDLRT 581

RESULT 14
CAN1_RABBIT STANDARD; PRT; 302 AA.
AC P06815;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Calpain 1, large [catalytic] subunit (EC 3.4.22.52) (Calcium-activated
DE neutral proteinase) (CANP) (Mu-type) (mucANP) (Micromolar-calpain)
DE (Fragment).
GN Name=CANP1;
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_taxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8625092; PubMed=2424911;
RA Emori Y., Kawasaki H., Sugihara H., Imajoh S., Kawashima S.,
RA Suzuki K.;
RA "Isolation and sequence analyses of cDNA clones for the large subunits
RT of two isozymes of rabbit calcium-dependent protease.";
RL J. Biol. Chem. 261:9465-9471(1986).
RN [2]
RP CALCIUM-BINDING DATA.
RX MEDLINE=8727982; PubMed=3038855;
RA Minami Y., Emori Y., Kawasaki H., Suzuki K.;
RA "3-F hand structure-domain of calcium-activated neutral protease
RT

```

```

RT (CANP) can bind Ca2+ ions.";
RL J. Biochem. 101:899-995(1987).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction.
CC -1- CATALYTIC ACTIVITY: Broad endopeptidase specificity.
CC -1- COFACTOR: Binds 3 calcium ions (By similarity).
CC -1- ENZYME REGULATION: Activated by micromolar concentrations of
CC calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Forms a heterodimer with a small (regulatory) subunit
CC (CAPNS1).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca(2+) binding (By similarity).
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- SIMILARITY: Belongs to peptidase family C2.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M13363; AAA31456.1; -
CC PIR; A24815; A24815.
CC HSP; P17655; LKFX.
CC MEROPS; C02.001; -.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR010983; EF_Hand_Like.
CC InterPro; IPR001300; Peptidase C2.
CC InterPro; IPR000169; Peptidase C2.
CC Pfam; PF01067; Calpain_III; 1.
CC Pfam; PF00036; ehand_3.
CC SMART; SM00720; Calpain_III; 1.
CC SMART; SM00054; EFh; 3.
CC PROSITE; PS00018; EF_HAND; 2.
CC PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.
CC PROSITE; PS00639; THIOL_PROTEASE_CYS; PARTIAL.
CC PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.
CC Calcium-binding; Hydrolase; Multigene family; Thiol protease.
KW NON_TER 1
FT DOMAIN <1 114 Domain III.
FT DOMAIN 115 130 Linker.
FT DOMAIN 131 301 Domain IV.
FT DOMAIN 142 153 Ancestral calcium site 1 (Potential).
FT CA_BIND 186 197 EF-hand 2.
FT CA_BIND 216 227 EF-hand 3.
FT DOMAIN 251 292 Ancestral calcium site 4 (Potential).
FT DOMAIN 281 292 Ancestral calcium site 5 (Potential).
SQ SEQUENCE 302 AA; 35274 MW; 178BFBF4216C6EAB CRC64;

Query Match 89.0%; Score 121; DB 1; Length 302;
Best Local Similarity 85.7%; Pred. No. 1e-07;
Matches 24; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHKDLRT 28
Db 143 AGEDLEISVRELQILNRIISKHKDLRT 170

RESULT 15
Q7SYA9 PRELIMINARY; PRT; 704 AA.
ID Q7SYA9
AC Q7SYA9;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein zgc:63788.
GN Name=zgc:63788;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB; TISSUE=Whole body;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kertman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB; TISSUE=Whole body;
 RA Strausberg R.;
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL, SC054941; RAH54941.1; -
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0004198; F:calpain activity; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR001300; Peptidase C2.
 DR InterPro; IPR000169; Pept_cys_acsite.
 DR Pfam; PF01067; Calpain_III; 1.
 DR Pfam; PF00036; efhand; 2.
 DR Pfam; PF00648; Peptidase_C2; 1.
 DR ProDom; PD000012; EF-hand; 1.
 DR SMART; SM00720; calpain_III; 1.
 DR SMART; SM00230; CysPc; 1.
 DR SMART; SM00054; EFh; 2.
 DR PROSITE; PS02023; CALPAIN_CAT; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS00139; THIOLEPROTEASE_CYS; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 704 AA; 80403 MW; EB68F95EF1AF273 CRC64;

Query Match 87.5%; Score 119; DB 2; Length 704;
 Best Local Similarity 78.6%; Pred. No. 4.4e-07;
 Matches 22; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGEDMEISVKELRTILNRIISKHDLRT 28
 Db 546 AGEDMEISVHELKTLNRIIVSKHDKVKT 573

Search completed: November 10, 2004, 12:27:20
 Job time : 32.6358 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 7.74383 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-19

Perfect score: 131

Sequence: 1 GLFEESEYMAAADEYNRLKQVKQPA 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB pep.*
5: /cgn2_6/ptodata/1/iaa/PCITUS COMB pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1 pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	80.9	108	3	US-09-026-343-8
2	106	80.9	108	3	US-09-362-871-8
3	106	80.9	521	4	US-08-945-826-6
4	106	80.9	521	4	US-09-197-503-6
5	106	80.9	522	3	US-09-142-732-2
6	106	80.9	522	4	US-08-945-826-2
7	106	80.9	522	4	US-08-137-503-2
8	105	80.2	521	4	US-08-945-826-4
9	105	80.2	521	4	US-09-197-503-4
10	51	38.9	488	3	US-09-248-796A-20028
11	50.5	38.5	676	4	US-09-085-199B-7
12	50.5	38.5	890	4	US-08-849-602-19
13	50	38.2	885	4	US-08-710-279-1660
14	50	38.2	1211	3	US-08-134-001C-4820
15	48.5	37.0	83	2	US-08-499-676A-10
16	48	36.6	551	4	US-09-252-991A-30116
17	47	35.9	323	2	US-08-540-804-18
18	47	35.9	323	2	US-08-218-265-18
19	47	35.9	323	3	US-08-521-872-18
20	47	35.9	323	3	US-08-530-399-18
21	47	35.9	323	3	US-08-538-092-651
22	47	35.9	878	3	US-08-941-936-2
23	46	35.1	63	4	US-09-107-532A-6771
24	46	35.1	386	4	US-08-338-352-4722
25	46	35.1	640	2	US-08-671-978A-10
26	45.5	34.7	1068	3	US-09-085-199B-11
27	45	34.4	207	4	US-09-252-991A-31948

28	45	34.4	244	4	US-09-134-000C-4656	Sequence 4656, Ap
29	45	34.4	257	4	US-09-248-796A-20922	Sequence 20922, A
30	45	34.4	867	4	US-09-248-796A-19207	Sequence 19207, A
31	45	34.4	890	3	US-09-342-648-10	Sequence 10, Appl
32	45	34.4	1112	4	US-09-717-364A-27	Sequence 27, Appl
33	44.5	34.0	623	4	US-08-714-741-47	Sequence 47, Appl
34	44	33.6	194	4	US-09-489-039A-12868	Sequence 12868, A
35	44	33.6	243	3	US-08-965-087-2	Sequence 2, Appl
36	44	33.6	331	4	US-09-270-767-41224	Sequence 41224, A
37	44	33.6	331	4	US-09-270-767-56440	Sequence 56440, A
38	44	33.6	340	3	US-08-960-780-34	Sequence 34, Appl
39	44	33.6	340	3	US-09-073-898-34	Sequence 34, Appl
40	44	33.6	340	4	US-09-307-106-10	Sequence 10, Appl
41	44	33.6	340	4	US-09-850-351A-34	Sequence 34, Appl
42	44	33.6	437	4	US-09-328-352-5193	Sequence 5193, Ap
43	44	33.6	514	4	US-09-678-686B-22	Sequence 22, Appl
44	44	33.6	551	4	US-09-198-452A-653	Sequence 653, App
45	44	33.6	552	3	US-08-557-006C-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-09-026-343-8
; Sequence 8, Application US/09026343
; Patent No. 6008018
; GENERAL INFORMATION:
; APPLICANT: DUAN, D. ROXANNE
; APPLICANT: SHILATIFARD, ALI
; APPLICANT: CONAWAY, JOAN W.
; APPLICANT: CONAWAY, RONALD C.
; TITLE OF INVENTION: ELL2, A New Member of an ELL Family of
; TITLE OF INVENTION: RNA Polymerase II Elongation Factors
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/026,343
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/038,447
; FILING DATE: 19-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDSTEIN, JORGE A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1488.0880001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 372-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-09-026-343-8

Query Match 80.9%; Score 106; DB 3; Length 108;
Best Local Similarity 91.7%; Pred. No. 4.6e-09;
Matches 22; Conservative 2; Indels 0; Gaps 0;

QY 3 REESEYMAADEVNRLKQVKQA 26
 Db 54 REESEYMAADEVNRLKQVKGA 77

RESULT 2

US-09-362-871-8
 ; Sequence 8, Application US/09362871
 ; Patent No. 6379923
 ; GENERAL INFORMATION:
 ; APPLICANT: DUAN, D. ROXANNE
 ; APPLICANT: SHILATIFARD, ALI
 ; APPLICANT: CONAWAY, JOAN W.
 ; APPLICANT: CONAWAY, RONALD C.
 ; TITLE OF INVENTION: ELL2, A New Member of an ELL Family of
 ; RNA Polymerase II Elongation Factors
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 ; STREET: 1100 NEW YORK AVENUE, SUITE 600
 ; CITY: WASHINGTON
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/362,871
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/026,343
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: GOLDSTEIN, JORGE A.
 ; REGISTRATION NUMBER: 29,021
 ; REFERENCE/DOCKET NUMBER: 1498.0880001
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 372-2540
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 108 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: protein
 ; US-09-362-871-8

Query Match 80.9%; Score 106; DB 3; Length 108;
 Best Local Similarity 91.7%; Pred. No. 4.6e-09;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 REESEYMAADEVNRLKQVKQA 26
 Db 54 REESEYMAADEVNRLKQVKGA 77

RESULT 3

US-08-945-826-6
 ; Sequence 6, Application US/08945826
 ; Patent No. 6489460
 ; GENERAL INFORMATION:
 ; APPLICANT: TSUKITA, Shoichiro
 ; TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
 ; STREET: P.O. Box 747

; CITY: Falls Church
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22040-0747
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/945,826
 ; FILING DATE: 05-NOV-1997
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: FARACI, C. Joseph
 ; REGISTRATION NUMBER: 32,350
 ; REFERENCE/DOCKET NUMBER: 0425-0660P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 205-8000
 ; TELEFAX: (703) 205-8050
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 521 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-945-826-6

Query Match 80.9%; Score 106; DB 4; Length 521;
 Best Local Similarity 91.7%; Pred. No. 2.9e-08;
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 REESEYMAADEVNRLKQVKQA 26
 Db 467 REESEYMAADEVNRLKQVKGA 490

RESULT 4

US-09-197-503-6
 ; Sequence 6, Application US/09197503
 ; Patent No. 6559286
 ; GENERAL INFORMATION:
 ; APPLICANT: TSUKITA, Shoichiro
 ; TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
 ; STREET: P.O. Box 747
 ; CITY: Falls Church
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22040-0747
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/197,503
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/945,826
 ; FILING DATE: 05-NOV-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: FARACI, C. Joseph
 ; REGISTRATION NUMBER: 32,350
 ; REFERENCE/DOCKET NUMBER: 0425-0660P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 205-8000
 ; TELEFAX: (703) 205-8050
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:


```

; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-197-503-6

Query Match      80.9%; Score 106; DB 4; Length 521;
Best Local Similarity 91.7%; Pred. No. 2.9e-08;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 REESEYMAADEYNRLKQVKQA 26
Db 467 REESEYMAADEYNRLKQVKGSA 490

RESULT 5
US-09-142-732-2
; Sequence 2, Application US/09142732
; Patent No. 6252045
; GENERAL INFORMATION:
; APPLICANT: James M. Anderson
; APPLICANT: Christina M. Van Itallie
; TITLE OF INVENTION: Human Occludin, Its Uses and
; TITLE OF INVENTION: Enhancement of Drug Absorption Using Occludin Inhibitors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University Medical School
; ADDRESSEE: Section of Digestive Diseases
; ADDRESSEE: Department of Internal Medicine
; STREET: 333 Cedar Street, LCI 105
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: United States of America
; ZIP: 06520-8057
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 1.44 Mb diskette
; COMPUTER: IBM PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Word Processing
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/142,732
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/05809
; FILING DATE: March 14, 1997
; APPLICATION NUMBER: U.S. 60/013,625
; FILING DATE: March 15, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary M. Krinsky
; REGISTRATION NUMBER: 32423
; REFERENCE/DOCKET NUMBER: 1751-P0016B.PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-324-6155
; TELEFAX: 203-327-1096
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: polypeptide
; FRAGMENT TYPE: complete sequence
; FEATURE:
; NAME/KEY: human occludin
; US-09-142-732-2

Query Match      80.9%; Score 106; DB 3; Length 522;
Best Local Similarity 91.7%; Pred. No. 2.9e-08;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 REESEYMAADEYNRLKQVKQA 26

```

```

Db 468 REESEYMAADEYNRLKQVKGSA 491

RESULT 6
US-08-945-826-2
; Sequence 2, Application US/08945826
; Patent No. 6489460
; GENERAL INFORMATION:
; APPLICANT: TSUKITA, Shoichiro
; TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/945,826
; FILING DATE: 05-NOV-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. Joseph
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 0425-0660P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-945-826-2

Query Match      80.9%; Score 106; DB 4; Length 522;
Best Local Similarity 91.7%; Pred. No. 2.9e-08;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 REESEYMAADEYNRLKQVKQA 26
Db 468 REESEYMAADEYNRLKQVKGSA 491

RESULT 7
US-09-197-503-2
; Sequence 2, Application US/09197503
; Patent No. 6559286
; GENERAL INFORMATION:
; APPLICANT: TSUKITA, Shoichiro
; TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

QY 3 REESEYMAADEVNRLKQVKA 26
Db 467 REESEYMAADEVNRLKQVK 487

US-09-197-503-2

Query Match 80.9%; Score 106; DB 4; Length 522;
Best Local Similarity 91.7%; Pred. No. 2.9e-08;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 8

US-08-945-826-4

Sequence 4, Application US/08945826
Patent No. 6489460

GENERAL INFORMATION:
APPLICANT: TSUKITA, Shoichiro
TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
STREET: P.O. Box 747
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22040-0747

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,826
FILING DATE: 05-NOV-1997

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/945,826
FILING DATE: 05-NOV-1997

ATTORNEY/AGENT INFORMATION:
NAME: FARACI, C. Joseph
REGISTRATION NUMBER: 32,350
REFERENCE/DOCKET NUMBER: 0425-0660P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 522 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

QY 3 REESEYMAADEVNRLKQVKA 26
Db 468 REESEYMAADEVNRLKQVKA 491

US-09-197-503-2

Query Match 80.9%; Score 106; DB 4; Length 522;
Best Local Similarity 91.7%; Pred. No. 2.9e-08;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 8

US-08-945-826-4

Sequence 4, Application US/08945826
Patent No. 6489460

GENERAL INFORMATION:
APPLICANT: TSUKITA, Shoichiro
TITLE OF INVENTION: HUMAN ADHESION MOLECULE OCCLUDIN
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
STREET: P.O. Box 747
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22040-0747

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,826
FILING DATE: 05-NOV-1997

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FARACI, C. Joseph
REGISTRATION NUMBER: 32,350
REFERENCE/DOCKET NUMBER: 0425-0660P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

QY 3 REESEYMAADEVNRLKQVKA 23
Db 467 REESEYMAADEVNRLKQVK 487

US-09-197-503-4

Query Match 80.2%; Score 105; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10

US-09-248-796A-20028

Sequence 20028, Application US/09248796A
Patent No. 6747137

GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 20028

QY 3 REESEYMAADEVNRLKQVKA 23
Db 467 REESEYMAADEVNRLKQVK 487

US-09-197-503-4

Query Match 80.2%; Score 105; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10

US-09-248-796A-20028

Sequence 20028, Application US/09248796A
Patent No. 6747137

GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 20028

```
US-09-849-602-19
; Sequence 19, Application US/09849602
; Patent No. 6794501
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Old, Lloyd J.
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Chen, Yao-Tsung
; FILE REFERENCE: L0461/7105(JRV)
; CURRENT APPLICATION NUMBER: US/09/849,602
; CURRENT FILING DATE: 2001-05-04
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-602-19

Query Match          38.5%; Score 50.5; DB 4; Length 890;
Best Local Similarity 50.0%; Pred.No.23;
Matches      11; Conservative    5; Mismatches     5; Indels       1; Gaps         1;

QY      1 GLRESEEEYMAAAD-EYNRLKQ 21
        |||||: |: |:: ||::
DB      253 GLREEAEAKASATEARYNKLE 274

RESULT 13
US-09-710-279-1660
; Sequence 1660, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480U5
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-1660

Query Match          38.2%; Score 50; DB 4; Length 885;
Best Local Similarity 50.0%; Pred.No.27;
Matches           8; Conservative   5; Mismatches    3; Indels       0; Gaps         0;

QY              6 SEEYMAAADEYNELKQ 21
                ||:||:|::|||||
DB             510 SEEYFSASQQNNLKE 525

RESULT 14
US-09-134-001C-4820
; Sequence 4820, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: gpc-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
```

; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4820
 ; LENGTH: 1211
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4820

Query Match 38.2%; Score 50; DB 3; Length 1211;
 Best Local Similarity 50.0%; Pred.No. 39;
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
 QY 6 SBEYMAADEYNRLKQ 21
 Db 734 SEQYFSASQYNNLKE 749

RESULT 15
 US-08-499-676A-10
 ; Sequence 10, Application US/08499676A
 ; Patent No. 5998154
 ; GENERAL INFORMATION:
 ; APPLICANT: AGNES SCHONBRUNN
 ; TITLE OF INVENTION: SOMATOSTATIN RECEPTOR PEPTIDE
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
 ; STREET: 1177 West Loop South, 10th Floor
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77027-9095
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/499,676A
 ; FILING DATE: July 7, 1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Krieger, Paul E.
 ; REGISTRATION NUMBER: 25,886
 ; REFERENCE/DOCKET NUMBER: 79247.3/A95175US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 713-850-0909
 ; TELEFAX: 713-850-0165
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 83 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-08-499-676A-10

Query Match 37.0%; Score 48.5; DB 2; Length 83;
 Best Local Similarity 50.0%; Pred.No. 2.9;
 Matches 12; Conservative 1; Mismatches 10; Indels 1; Gaps 1;
 QY 3 REESEYMAADEYN-RLKQVKQP 25
 Db 49 REEEERMQGQEMNRLSQAQP 72

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 15:53:52 ; Search time 25.0772 Seconds
(without alignments)
366.225 Million cell updates/sec

US-10-092-750-19

Title:
Perfect score: 131

Sequence: 1 GLREESEYMAADEVNRLKQVKQA 26

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:

1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	131	100.0	26	14 US-10-092-750-19
2	106	80.9	108	13 US-10-028-780-8
3	106	80.9	522	9 US-09-891-064A-2
4	106	80.9	522	14 US-10-176-847-104
5	106	80.9	522	15 US-10-501-953-900
6	55	42.0	175	14 US-10-094-749-1982
7	53	40.5	330	14 US-10-369-493-18880
8	52	39.7	330	14 US-10-369-493-19971
9	52	39.7	481	9 US-09-748-875-6
10	52	39.7	481	10 US-09-298-523B-6
11	52	39.7	496	9 US-09-748-875-4
12	52	39.7	496	9 US-09-298-523B-4
13	52	39.7	1163	15 US-10-282-122A-51864

14 51 38.9 1126 14 US-10-369-493-13544 Sequence 13544, A
15 50.5 38.5 110 10 US-09-882-227-8 Sequence 8, Appli
16 50.5 38.5 890 10 US-09-849-602-19 Sequence 19, Appli
17 50 38.2 305 17 US-10-425-115-286036 Sequence 286036, A
18 50 38.2 1189 15 US-10-282-122A-70320 Sequence 70320, A
19 50 38.2 1581 17 US-10-425-115-199277 Sequence 199277, A
20 49.5 37.8 152 16 US-10-767-701-54861 Sequence 54861, A
21 49.5 37.8 536 15 US-10-425-114-70258 Sequence 70258, A
22 49.5 37.8 617 17 US-10-739-930-10810 Sequence 10810, A
23 49.5 37.8 618 17 US-10-425-115-266685 Sequence 266685, A
24 49.5 37.8 634 15 US-10-425-114-45740 Sequence 45740, A
25 49 37.4 183 17 US-10-425-115-286031 Sequence 286031, A
26 49 37.4 255 15 US-10-425-114-47386 Sequence 47386, A
27 49 37.4 285 15 US-10-425-114-58323 Sequence 58323, A
28 49 37.4 299 17 US-10-425-115-240819 Sequence 240819, A
29 49 37.4 299 17 US-10-425-115-286037 Sequence 286037, A
30 49 37.4 300 15 US-10-425-114-53048 Sequence 53048, A
31 49 37.4 300 15 US-10-425-115-286034 Sequence 286034, A
32 49 37.4 307 17 US-10-425-115-286035 Sequence 286035, A
33 49 37.4 307 17 US-10-425-115-286038 Sequence 286038, A
34 49 37.4 308 15 US-10-425-114-37032 Sequence 37032, A
35 49 37.4 308 15 US-10-425-114-39618 Sequence 39618, A
36 49 37.4 308 15 US-10-425-114-60362 Sequence 60362, A
37 49 37.4 308 15 US-10-425-114-63528 Sequence 63528, A
38 49 37.4 311 14 US-10-156-761-9698 Sequence 9698, Ap
39 37.4 359 16 US-10-437-963-158617 Sequence 158617, A
40 49 37.4 1465 15 US-10-320-797-3357 Sequence 3357, Ap
41 49 37.4 1948 14 US-10-032-585-7611 Sequence 7611, Ap
42 49 37.4 426 14 US-10-262-313-16 Sequence 16, Appli
43 48.5 37.0 426 16 US-10-768-878-16 Sequence 16, Appli
44 48.5 37.0 428 10 US-09-992-331-16 Sequence 16, Appli

ALIGNMENTS

RESULT 1

US-10-092-750-19
; Sequence 19, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Walpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-XI
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-19

Query Match 100.0%; Score 131; DB 14; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.3e-11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLREESEYMAADEVNRLKQVKQA 26

Db 1 GLREESEYMAADEVNRLKQVKQA 26

RESULT 2

US-10-028-780-8
; Sequence 8, Application US/10028780
; Publication No. US20020132329A1
; GENERAL INFORMATION:

APPLICANT: DUAN, D. ROXANNE
SHILATIFARD, ALI
CONAWAY, JOAN W.
CONAWAY, RONALD C.
TITLE OF INVENTION: ELL2, A New Member of an ELL Family of
RNA Polymerase II Elongation Factors
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/028,780
FILING DATE: 28-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/026,343
FILING DATE: <Unknown>
APPLICATION NUMBER: US 60/038,447
FILING DATE: 19-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: GOLDSTEIN, JORGE A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 1488.0880001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 372-2540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS: No. US20020132329A1 Relevant
TOPOLOGY: No. US20020132329A1 Relevant
MOLECULE TYPE: Protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-028-780-8

Query Match 80.9%; Score 106; DB 13; Length 108;
Best Local Similarity 91.7%; Pred. No. 3.5e-07;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 REESEYMAADEVNRLKQVKQPA 26
Db 54 REESEYMAADEVNRLKQVKGSA 77

RESULT 3
US-09-891-064A-2
Sequence 2, Application US/09891064A
Patent No. US20020082391A1
GENERAL INFORMATION:
APPLICANT: James M. Anderson
TITLE OF INVENTION: Human Occludin, Its Uses and Enhancement of Drug
Absorption Using Occludin Inhibitors
FILE REFERENCE: OCR-754.CIP
CURRENT APPLICATION NUMBER: US/09/891.064A
PRIOR FILING DATE: 2001-06-25
PRIOR APPLICATION NUMBER: US 09/142,732
NUMBER OF SEQ ID NOS: 6
SOFTWARE: MS DOS
SEQ ID NO 2
LENGTH: 522
TYPE: PRT

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: peptide
LOCATION: complete sequence
OTHER INFORMATION: human occludin
US-09-891-064A-2

Query Match 80.9%; Score 106; DB 9; Length 522;
Best Local Similarity 91.7%; Pred. No. 2e-06;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 REESEYMAADEVNRLKQVKQPA 26
Db 468 REESEYMAADEVNRLKQVKGSA 491

RESULT 4
US-10-176-847-104
Sequence 104, Application US/10176847
Publication No. US20030068636A1
GENERAL INFORMATION:
APPLICANT: Velby, Petter Ole
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
CANCER
FILE REFERENCE: MRI-039
CURRENT APPLICATION NUMBER: US/10/176,847
CURRENT FILING DATE: 2002-06-21
NUMBER OF SEQ ID NOS: 112
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 104
LENGTH: 522
TYPE: PRT
ORGANISM: Homo sapiens
US-10-176-847-104

Query Match 80.9%; Score 106; DB 14; Length 522;
Best Local Similarity 91.7%; Pred. No. 2e-06;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 REESEYMAADEVNRLKQVKQPA 26
Db 468 REESEYMAADEVNRLKQVKGSA 491

RESULT 5
US-10-601-953-900
Sequence 900, Application US/10601953
Publication No. US2004007540A1
GENERAL INFORMATION:
APPLICANT: Quay, Steven C.
TITLE OF INVENTION: Compositions And Methods For Modulating Physiology Of Epithelial
Junctional Adhesion Molecules For Enhanced Mucosal Delivery Of
Therapeutic Compounds
FILE REFERENCE: 02-03US
CURRENT APPLICATION NUMBER: US/10/601,953
CURRENT FILING DATE: 2003-06-24
PRIOR APPLICATION NUMBER: 60/392,512
PRIOR FILING DATE: 2002-06-28
NUMBER OF SEQ ID NOS: 900
SOFTWARE: PatentIn version 3.2
SEQ ID NO 900
LENGTH: 522
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-10-601-953-900

Query Match 80.9%; Score 106; DB 15; Length 522;
Best Local Similarity 91.7%; Pred. No. 2e-06;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;


```
Db      446 QOAEEDYARSEEEYNRLTQQPP 468
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-4
Query Match      39.7%; Score 52; DB 10; Length 496;
Best Local Similarity 39.1%; Pred. No. 72;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
QY      3 RESEYMAADEVNRLKQVKP 25
      ::|||::|||::|||::|||::|||
Db      461 QOAEEDYARSEEEYNRLTQQPP 483
;
;
;
RESULT 13
US-10-282-122A-51864
; Sequence 51864, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51864
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51864
Query Match      39.7%; Score 52; DB 15; Length 1163;
Best Local Similarity 52.6%; Pred. No. 1.8e-02;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY      4 RESEYMAADEVNRLKQV 22
      ::|||::|||::|||::|||::|||
Db      230 EASKEYLKAESEYNEGKEV 248
;
;
;
RESULT 11
US-09-748-875-4
; Sequence 4, Application US/09748875
; Publication No. US20010016200A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/748,875
; CURRENT FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: 09/298,523
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-748-875-4
Query Match      39.7%; Score 52; DB 9; Length 496;
Best Local Similarity 39.1%; Pred. No. 72;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
QY      3 RESEYMAADEVNRLKQVKP 25
      ::|||::|||::|||::|||::|||
Db      461 QOAEEDYARSEEEYNRLTQQPP 483
;
;
;
RESULT 12
US-09-298-523B-4
; Sequence 4, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
```



```

RESULT 14
US-10-369-493-13544
; Sequence 13544, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13544
; LENGTH: 1126
; TYPE: PRT
; ORGANISM: Thermoplasma volcanium
US-10-369-493-13544

Query Match          38.9%; Score 51; DB 14; Length 1126;
Best Local Similarity 69.2%; Pred.No. 2.4e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 7 EYMAAADEYNRL 19
    ||::|||::|||::|
DB 352 EFIAADEYNKL 364

RESULT 15
US-09-882-227-8
; Sequence 8, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-8

Query Match          38.5%; Score 50.5; DB 10; Length 110;
Best Local Similarity 37.5%; Pred. No. 23;
Matches 9; Conservative 9; Mismatches 5; Indels 1; Gaps 1;

QY 2 LRESEBYMAADBY-NRLQVKQ 24
DB 81 LKORNSEYKGESDKYKRIKELEQ 104

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:36:51 ; Search time 5.25617 Seconds
(without alignments)
475.942 Million cell updates/sec

Title: US-10-092-750-19
Perfect score: 131
Sequence: 1 GLREESEYMAAADEYNRLKQVKQA 26

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	80.9	522	2	G02533
2	59	45.0	504	2	A49467
3	55	42.0	764	2	T05409
4	53.5	40.8	110	2	A71921
5	53	40.5	330	2	AE2411
6	53	40.5	501	2	A55749
7	52	39.7	1163	2	G97236
8	51.5	39.3	300	2	H75200
9	51.5	39.3	300	2	D71234
10	51	38.9	339	2	S77001
11	50.5	38.5	110	2	E64591
12	49	37.4	406	2	T24492
13	48.5	37.0	428	2	S30508
14	48	36.6	188	2	C97142
15	48	36.6	228	2	D85045
16	48	36.6	736	1	VPXRPC
17	48	36.6	744	2	S45061
18	48	36.6	744	2	S45060
19	48	36.6	859	2	S49046
20	48	36.6	1854	2	F86378
21	47	35.9	648	2	UC5973
22	47	35.9	648	2	JC4674
23	47	35.9	878	2	T23537
24	46	35.1	184	1	LWLVI
25	46	35.1	221	2	D69487
26	46	35.1	378	2	A96982
27	46	35.1	582	2	F90603
28	46	35.1	2025	2	T21588
29	46	35.1	2561	2	T24864

hypothetical prote
L-arabinose isomer
probable acetyltra
hypothetical prote
hypothetical prote
hypothetical prote
TPR-repeat-contain
conserved hypothet
hypothetical prote
probable cell divi
probable protein k
hypothetical prote
long-chain-fatty-a
proline dehydrogen
malate dehydrogena
hypothetical prote

ALIGNMENTS

RESULT 1
G02533
occludin - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C:Accession: G02533
R:Van Itallie, C.M.
submitted to the EMBL Data Library, April 1996
A:Reference number: H01403
A:Accession: G02533
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-522 <VAN>
A:Cross-references: UNIPROT:Q16625; EMBL:U53823; NID:G1322281; PIDN:AAB00195.1; PID:G1332;
C:Superfamily: occludin

Query Match 80.9%; Score 106; DB 2; Length 522;
Best Local Similarity 91.7%; Pred No. 8.3e-07;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 REESEYMAAADEYNRLKQVKQA 26
DB 468 REESEYMAAADEYNRLKQVKQA 491

RESULT 2
A49467
occludin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C:Accession: A49467
R:Furuse, M.; Hirase, T.; Itoh, M.; Nagafuchi, A.; Yonemura, S.; Tsukita, S.; Tsukita, S.
J. Cell Biol. 123, 1777-1788, 1993
A>Title: Occludin: a novel integral membrane protein localizing at tight junctions.
A:Reference number: A49467; MUID:94103332; PMID:8276896
A:Accession: A49467
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-504 <FUR>
A:Cross-references: UNIPROT:Q91049; GB:D21837; NID:G464148; PIDN:BA04865.1; PID:G540494
C:Superfamily: occludin
C:Keywords: liver; membrane protein.

Query Match 45.0%; Score 59; DB 2; Length 504;
Best Local Similarity 43.5%; Pred. No. 2.3;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 2 LREESEYMAAADEYNRLKQVKQA 24
DB 449 ITSDSPQYQVAAEYNQLDKLR 471

RESULT 3

T05409
 hypothetical protein F10M6.170 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
 C:Accession: T05409
 R:Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzi, A.; N
 submitted to the Protein Sequence Database, February 1998
 A:Reference number: Z15414
 A:Accession: T05409
 A:Molecule type: DNA
 A:Residues: 1-764 <BEV>
 A:Cross-references: UNIPROT:O49371; EMBL:AL021811
 A:Experimental source: cultivar Columbia; BAC clone F10M6
 C:Genetics:
 A:Map position: 4
 A:Note: F10M6.170

Query Match 42.0%; Score 55; DB 2; Length 764;
 Best Local Similarity 45.0%; Pred. No. 13;
 Matches 9; Conservative 8; Mismatches 3; Indels 0; Gaps 0;
 QY 2 LRESEYMAAADEVNRLKQ 21
 Db 270 LKQTEWLIQAQDEVNKLKE 289

RESULT 4

A71921
 hypothetical protein jhp0520 - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
 C:Accession: A71921
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Malls, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:99120557; PMID:9923682
 A:Accession: A71921
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-110 <ARN>
 A:Cross-references: UNIPROT:Q9ZLQ8; GB:AE001485; GB:AE001439; NID:G4155057; PIDN:AXD0610
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp0520

Query Match 40.8%; Score 53.5; DB 2; Length 110;
 Best Local Similarity 41.7%; Pred. No. 2.6;
 Matches 10; Conservative 8; Mismatches 5; Indels 1; Gaps 1;
 QY 2 LRESEYMAAADEVNRLKQVKQ 24
 Db 81 LKQNSEYKGSCKYNNIKLEQ 104

RESULT 5

AE2411
 phenylalanine-tRNA synthetase alpha chain [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Ncte: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C:Accession: AE2411
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AE2411
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-330 <KUR>

A:Cross-references: UNIPROT:Q8YMT4; GB:BA000019; PIDN:BA076544.1; PID:G17133982; GSPDB:G
 A:Experimental source: strain PCC 7120

C:Genetics:
 A:Gene: all4845
 C:Superfamily: phenylalanine-tRNA ligase alpha chain

Query Match 40.5%; Score 53; DB 2; Length 330;
 Best Local Similarity 45.5%; Pred. No. 9.8;
 Matches 10; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LRESEYMAAADEVNRLKQVK 23

Db 12 LRQEGEQALAAADTLERLEELR 33

RESULT 6

A55749
 spliceosome-associated protein SAP 61 - human
 C:Species: Homo sapiens (man)
 C>Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 09-Jul-2004
 C:Accession: A55749; S53583; S49319
 R:Chirara, M.D.; Champion-Arnaud, P.; Buvoli, M.; Nadal-Ginard, B.; Reed, R.
 Proc. Natl. Acad. Sci. U.S.A. 91, 6403-6407, 1994
 A>Title: Specific protein-protein interactions between the essential mammalian spliceosom
 A:Reference number: A55749; MUID:94294390; PMID:8022796
 A:Accession: A55749
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-501 <CHI>
 A:Cross-references: UNIPROT:Q12874; GB:U08915; NID:G508722; PIDN:AAA19625.1; PID:G508723
 R:Kraemer, A.; Legrain, P.; Mulhauser, F.; Groening, K.; Broel, R.; Bilbe, G.
 Nucleic Acids Res. 22, 5223-5228, 1994
 A>Title: Splicing factor SF3a60 is the mammalian homologue of PRP9 of S. cerevisiae: the
 A:Reference number: S53583; MUID:95116308; PMID:7816610
 A:Accession: S53583
 A>Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-175; G, 177-501 <KRA>
 A:Cross-references: EMBL:X81789; NID:G551449; PIDN:CAA57388.1; PID:G551450

Query Match 40.5%; Score 53; DB 2; Length 501;
 Best Local Similarity 42.9%; Pred. No. 15;
 Matches 12; Conservative 6; Mismatches 6; Indels 4; Gaps 1;
 QY 1 GLRSESEYMAAADE----YNRLKQVKQ 24
 Db 66 GLRKEELNAISGPNEFAEFYNRLKQIKE 93

RESULT 7
 G97236
 Apase involved in DNA repair [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
 C:Accession: G97236
 R:Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J. Bacteriol. 183, 4823-4838, 2001
 A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: G97236
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1163 <KUR>
 A:Cross-references: UNIPROT:Q97FK3; GB:AE001437; PIDN:AAK0682.1; PID:G15025772; GSPDB:G
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC2736

Query Match 39.7%; Score 52; DB 2; Length 1163;
 Best Local Similarity 52.6%; Pred. No. 51;
 Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

RESULT 10
S77001
hypothetical protein slr0787 - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 16-Aug-2004
C:Accession: S77001
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3. 109-136. 1996

Query Match	37.4%	Score 49;	DB 2;	Length 406;
Best Local Similarity	37.5%	Pred. No. 43;		
Matches	9;	Mismatches	8;	Indels
	Conservative			0;
				Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01 ; Search time 28.4475 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-19

Perfect score: 131

Sequence: 1 GLREESEYMAADEYNRLKQVKQA 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	106	80.9	121	2	O97766	O97766 sus scrofa
2	106	80.9	449	2	O6PP84	O6PP84 bos taurus
3	106	80.9	449	2	RAT00455	RAT00455 bos taurus
4	106	80.9	521	1	OCLN_MOUSE	O61146 mus musculus
5	106	80.9	521	2	BAC31298	BAC31298 mus musculus
6	106	80.9	522	1	OCLN_HUMAN	Q16625 homo sapien
7	106	80.9	523	2	O6P6T5	O6P6T5 rattus norv
8	106	80.9	523	2	O92303	O92303 rattus norv
9	106	80.9	523	2	AAH62037	AAH62037 rattus norv
10	105	80.2	521	1	OCLN_CANFA	Q28259 canis fami
11	105	80.2	560	2	Q9N0W3	Q9N0W3 canis fami
12	79	60.3	493	1	OCLN_XENLA	Q9PUN1 xenopus lae
13	64	48.9	497	2	O6NRR0	O6NRR0 xenopus lae
14	64	48.9	497	2	AAH70670	AAH70670 xenopus l
15	59	45.0	504	1	OCLN_CHICK	O91049 gallus gall
16	58	44.3	489	1	OCLN_POTTR	O28793 pectorous tr
17	56	42.7	500	2	Q6NX39	Q6NX39 brachydania
18	56	42.7	500	2	AAH67178	AAH67178 brachydania
19	55	42.0	259	2	O96AA1	O96AA1 homo sapien
20	55	42.0	356	2	O7ZUC2	O7ZUC2 brachydania
21	55	42.0	356	2	AAH66581	AAH66581 brachydania
22	55	42.0	361	1	CB45_MOUSE	O61112 mus musculus
23	55	42.0	361	1	CB45_RAT	O31243 rattus norv
24	55	42.0	361	1	AAH68152	AAH68152 mus muscu
25	55	42.0	362	1	CB45_HUMAN	Q9BRK5 homo sapien
26	55	42.0	492	2	O7ZUE7	O7ZUE7 brachydania
27	55	42.0	764	2	O49371	O49371 arabidopsis
28	55	42.0	783	2	Q8H1E5	Q8H1E5 arabidopsis
29	55	42.0	783	2	O9C5L5	O9C5L5 arabidopsis
30	53.5	40.8	110	2	O92L08	O92L08 helicobacte
31	53	40.5	230	2	Q6TUF1	Q6TUF1 rattus norv

32	53	40.5	230	2	AAQ91049	AAQ91049 rattus no
33	53	40.5	330	1	SYFA_ANASP	Q8YMT4 anabaena sp
34	53	40.5	501	1	S3A3_HUMAN	Q12874 homo sapien
35	53	40.5	501	1	S3A3_MOUSE	Q94554 mus musculu
36	53	40.5	501	2	BAC27111	BAC27111 mus muscu
37	52	39.7	827	2	Q7VG33	Q7VG33 helicobacte
38	52	39.7	988	2	Q8GQW4	Q8GQW4 acinetobact
39	52	39.7	1163	1	SBCC_CLOAB	Q97FK1 clostridium
40	51.5	39.3	300	2	O57874	O57874 pyrococcus
41	51.5	39.3	300	2	Q9V2E3	Q9V2E3 pyrococcus
42	51	38.9	219	2	Q948F7	Q948F7 oryza sativ
43	51	38.9	300	2	Q8U4L2	Q8U4L2 pyrococcus
44	51	38.9	325	2	Q6DHE7	Q6DHE7 brachydania
45	51	38.9	339	1	NADM_SYNY3	Q55928 s bifunctio

ALIGNMENTS

RESULT 1

O97766	PRELIMINARY;	PRT;	121 AA.
ID	O97766		
AC	O97766;		
DT	01-MAY-1999 (TREMELrel. 10, Created)		
DT	01-MAY-1999 (TREMELrel. 10, Last sequence update)		
DT	01-MAR-2004 (TREMELrel. 26, Last annotation update)		
DE	Occludin (Fragment).		
OS	Sus scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
OX	NCBI_TaxID=9823;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97388360; PubMed=9247194;		
RA	Hirase T., Fujimoto K., Saitou M., Ando-Akatsuka Y., Itoh M.,		
RA	Furuse M., Tsukita S., Rubin L.L.;		
RT	"Occludin as a possible determinant of tight junction permeability in endothelial cells."		
RL	J. Cell Sci. 110:1603-1613 (1997).		
DR	EMBL; U79554; AAC00349.1; -		
DR	InterPro; IPR010844; Occludin_ELL.		
DR	Pfam; PF07303; Occludin_ELL; 1.		
FT	NON_TER 1		
FT	NON_TER 121		
SQ	SEQUENCE 121 AA; 14433 MW; 6B37FD57A5F647A6 CRC64;		

Query Match 80.9%; Score 106; DB 2; Length 121;
Best Local Similarity 91.7%; Pred. No. 8.1e-07;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 REESEYMAADEYNRLKQVKQA 26

Db 71 REESEYMAADEYNRLKQVKQA 94

RESULT 2

O6PP84	PRELIMINARY;	PRT;	449 AA.
ID	O6PP84		
AC	O6PP84;		
DT	05-JUL-2004 (TREMELrel. 27, Created)		
DT	05-JUL-2004 (TREMELrel. 27, Last sequence update)		
DT	05-JUL-2004 (TREMELrel. 27, Last annotation update)		
DE	Occludin (Fragment).		
GN	Names=OCLN;		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Retinal endothelium;		
RA	Sundstrom J.M., Felinski E.A., Phillips B., Antonetti D.A.;		


```

DR PRINTS; PR01258; OCCLUDIN
KW Coiled coil; Phosphorylation; Tight junction; Transmembrane.
FT DOMAIN 1 66 Cytoplasmic (Potential).
FT TRANSMEM 67 89 Potential.
FT DOMAIN 90 133 Extracellular (Potential).
FT TRANSMEM 134 158 Potential.
FT DOMAIN 159 168 Cytoplasmic (Potential).
FT TRANSMEM 169 193 Potential.
FT DOMAIN 194 241 Extracellular (Potential).
FT TRANSMEM 242 263 Potential.
FT DOMAIN 264 521 Cytoplasmic (Potential).
FT TRANSMEM 521 590 MW; 21D62F370EB33E7D CRC64;
FT DOMAIN 307 310 Poly-Pro.
FT TRANSMEM 310 424 TyR/GLY-RICH.
FT DOMAIN 424 488 Coiled coil (Potential).
FT TRANSMEM 488 521 AA; 59000 MW; 21D62F370EB33E7D CRC64;
SQ SEQUENCE 521 AA; 59000 MW; 21D62F370EB33E7D CRC64;

Query Match 80.9%; Score 106; DB 1; Length 521;
Best Local Similarity 91.7%; Pred. No. 3.6e-06;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 REESEYMAAADYVNLKQVKQA 26
Db 467 REESEYMAAADYVNLKQVKGSA 490

RESULT 5
BAC31298 PRELIMINARY; PRT; 521 AA.
AC BAC31298;
DT 14-APR-2004 (T-EMBLrel. 27, Created)
DT 14-APR-2004 (T-EMBLrel. 27, Last sequence update)
DT 14-APR-2004 (T-EMBLrel. 27, Last annotation update)
DE 7 days neonate cerebellum cDNA, RIKEN full-length enriched library,
DE clone: A730007CJ9 product: occludin, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;

```

```

RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kasinwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multipillar sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Horii F., Imciani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Saeki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK042576; BAC31298.1;
SQ SEQUENCE 521 AA; 58999 MW; 21D62F370EB33E7D CRC64;

Query Match 80.9%; Score 106; DB 2; Length 521;
Best Local Similarity 91.7%; Pred. No. 3.6e-06;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 REESEYMAAADYVNLKQVKQA 26
Db 467 REESEYMAAADYVNLKQVKGSA 490

RESULT 6
OCIN HUMAN
ID OCIN HUMAN STANDARD; PRT; 522 AA.
AC Q16625; Q8N6K1;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Occludin.
GN Name=OCIN;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RX MEDLINE=96181088; PubMed=8601611;
RA Ando-Akatsuka Y., Saitou M., Hirase T., Kishi M., Sakakibara A.,
RA Itoh M., Yonemura S., Furuse M., Tsukita S.;
RT "Interspecies diversity of the occludin sequence: cDNA cloning of
RT human, mouse, dog, and rat-kangaroo homologues.";
RL J. Cell Biol. 133:43-47(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA van Itallie C.M., Fanning A.S., Anderson J.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Fukasawa M., Toyota T., Yoshitsu K., Yoshikawa T.;
RT "Genomic structure of occludin gene.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;

```

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May play a role in the formation and regulation of the
CC tight junction (TJ) paracellular permeability barrier.
CC -!- SUBUNIT: Interacts with TUP1/ZO1 and with VAPA.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Localized at tight junctions of both
CC epithelial and endothelial cells. Highly expressed in kidney. Not
CC detected in testis.
CC -!- DOMAIN: The C-terminal is cytoplasmic and is important for
CC interaction with ZO-1. Sufficient for the tight junction
CC localization. Involved in the regulation of the permeability
CC barrier function of the tight junction (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the ELL / Occludin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: U49184; AAC50451.1; -;
CC EMBL: U53823; AAB00195.1; -;
CC EMBL: AF400630; AAL47094.1; -;
CC EMBL: AF400623; AAL47094.1; JOINED.
CC EMBL: AF400624; AAL47094.1; JOINED.
CC EMBL: AF400625; AAL47094.1; JOINED.
CC EMBL: AF400626; AAL47094.1; JOINED.
CC EMBL: AF400627; AAL47094.1; JOINED.
CC EMBL: AF400628; AAL47094.1; JOINED.
CC EMBL: AF400629; AAL47094.1; JOINED.
CC EMBL: BC029886; AAR29886.1; -;
CC PIR: G02533; G02533.
CC Genew: HGNC:8104; OCLN.
CC
CC GO: GO:0005886; C:plasma membrane; TAS.
CC GO: GO:0006461; P:protein complex assembly; TAS.
CC InterPro: IPR008253; Marvel.
CC InterPro: IPR002958; Occludin.
CC InterPro: IPR010844; Occludin_ELL.
CC Pfam: PF01284; MARVEL; 1.
CC Pfam: PF07303; Occludin_ELL; 1.
CC PRINTS: PR01258; OCLUDIN.
CC Coiled coil; Phosphorylation; Tight junction; Transmembrane.
CC DOMAIN 1 66 Cytoplasmic (Potential).
CC TRANSMEM 67 89 Potent.
CC DOMAIN 90 135 Extracellular.
CC TRANSMEM 136 160 Potent.
CC DOMAIN 161 170 Cytoplasmic (Potential).
CC TRANSMEM 171 195 Potent.
CC DOMAIN 196 243 Extracellular (Potential).
CC FT

FT TRANSMEM 244 255 Potential.
FT DOMAIN 256 522 Cytoplasmic (Potential).
FT TYR/GLY-RICH 131
FT DOMAIN 426 489 Coiled coil (Potential).
FT CONFLICT 233 233 L -> S (in Ref. 4).
FT SEQUENCE 522 AA; 59143 MW; AOCF9574BCF6E974 CRC64;
Query Match 80.9%; Score 106; DB 1; Length 522;
Best Local Similarity 91.7%; Pred. No. 3.6e-06;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 REESEYMAADEYNRLKQVKQPA 26
|||||
DB 468 REESEYMAADEYNRLKQVKQSA 491
|||||
RESULT 7
Q6P6T5 PRELIMINARY; PRT; 523 AA.
AC Q6P6T5
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DE Occludin.
GN Name=Occln;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
OX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strausberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC062037; AAH62037.1; -;
DR InterPro: IPR008253; Marvel.
DR InterPro: IPR002958; Occludin.
DR InterPro: IPR010844; Occludin_ELL.
DR Pfam: PF01284; MARVEL; 1.
DR Pfam: PF07303; Occludin_ELL; 1.
DR PRINTS: PR01258; OCLUDIN.
SQ SEQUENCE 523 AA; 59186 MW; A31E007B58AE5B5 CRC64;
Query Match 80.9%; Score 106; DB 2; Length 523;
Best Local Similarity 91.7%; Pred. No. 3.6e-06;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 REESEYMAADEYNRLKQVKQPA 26
|||||

```
Db 469 REESEYMAAADEYNRLKQVKGSA 492

RESULT 8
Q92303 PRELIMINARY; PRT; 523 AA.
AC Q92303
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Occludin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SD; TISSUE=Liver;
RA Kokai Y., Kuwahara K., Atsumi S., Mori M.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016425; BAA36681.1;
DR GO; GO:0016020; C-membrane; IEA.
DR InterPro; IPR008253; Marvel.
DR InterPro; IPR002958; Occludin.
DR InterPro; IPR010844; Occludin_ELL.
DR Pfam; PF01284; MARVEL; 1.
DR Pfam; PF07303; Occludin_ELL; 1.
DR PRINTS; PR01258; OCCLUDIN.
SQ SEQUENCE 523 AA; 59214 MW; 5E1E007568AE5B1 CRC64;

Query Match 80.9%; Score 106; DB 2; Length 523;
Best Local Similarity 91.7%; Pred. No. 3.6e-06;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 REESEYMAAADEYNRLKQVKGSA 26
|||||
Db 469 REESEYMAAADEYNRLKQVKGSA 492

RESULT 9
AAH62037 PRELIMINARY; PRT; 523 AA.
AC AAH62037;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE Occludin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.W., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Shapletton M., Soares M.B., Boraudo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalusz D.E., Schnurch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

Db 469 REESEYMAAADEYNRLKQVKGSA 492

QY 3 REESEYMAAADEYNRLKQVKGSA 26
|||||
Db 469 REESEYMAAADEYNRLKQVKGSA 492

RESULT 10
OCLN_CANFA STANDARD; PRT; 521 AA.
AC Q28269;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Occludin.
GN Name=OCLN;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=96181088; PubMed=8601611;
RA Ando-Akatsuka Y., Saitou M., Hirase T., Kishi M., Sakakibara A.,
RA Itoh M., Yonemura S., Furuse M., Tsukita S.;
RT "Interspecies diversity of the occludin sequence: cDNA cloning of
RT human, mouse, dog, and rat-kangaroo homologues.";
RL J. Cell Biol. 133:43-47(1996).
RN [2]
RP PHOSPHORYLATION.
RX MEDLINE=97327764; PubMed=9182670;
RA Sakakibara A., Furuse M., Saitou M., Ando-Akatsuka Y., Tsukita S.;
RT "Possible involvement of phosphorylation of occludin in tight junction
RT formation.";
RL J. Cell Biol. 137:1393-1401(1997).
CC -!- FUNCTION: May play a role in the formation and regulation of the
CC tight junction (TJ) paracellular permeability barrier. Interacts
CC with ZO-1.
CC -!- SUBUNIT: Interacts with VAPA (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Localized at tight junctions of both
CC epithelial and endothelial cells.
CC -!- DOMAIN: The C-terminal is cytoplasmic and is important for
CC interaction with ZO-1. Necessary for the tight junction
CC localization. Involved in the regulation of the permeability
CC barrier function of the tight junction (By similarity).
CC -!- PTM: Phosphorylated. Less-phosphorylated forms are found in
CC basolateral membrane, cytosol and tight junction. More-heavily
CC phosphorylated forms are concentrated exclusively in tight
CC junction.
CC -!- SIMILARITY: Belongs to the ELL / occludin family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

1100